THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 1 January 1990

Contents

(ISSN 0002-9890)

ARTICLES

Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to Leon Henkin HENRY ALDER AND IVAN NIVEN 1

The Bitangent Sphere Problem PETER J. GIBLIN AND DONAL B. O'SHEA 5

The σ-Game and Cellular Automata KLAUS SUTNER 24

ACKNOWLEDGEMENTS 35

UNSOLVED PROBLEMS

More on Mortality MELVEN KROM AND MYREN KROM 37

NOTES

Elementary Evaluations of $\int_0^\infty e^{-x^2}\,dx$, $\int_0^\infty \cos x^2\,dx$, and $\int_0^\infty \sin x^2\,dx$ ROBERT WEINSTOCK 39

The L^4 Norm of a Polynomial with Coefficients ± 1 DONALD J. NEWMAN AND J. S. BYRNES 42

Isoptic Characterization of a Circle (Proof of a Conjecture of M. S. Klamkin) JOHANNES C. C. NITSCHE 45

Two Theorems on Singular Values and Eigenvalues ROY MATHIAS 47

THE TEACHING OF MATHEMATICS

A Decomposition Theorem for Finite Automata TERESA L. SANTA COLOMA AND RALPH P. TUCCI 51

A Counterexample in Galois Theory THOMAS W. HUNGERFORD 54

A Simple Algorithm for Finding Eigenvalues and Eigenvectors for 2×2 Matrices TYRE A. NEWTON 57

Magic Squares and Linear Algebra ARNO VAN DEN ESSEN 60

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 63

Advanced Problems and Solutions 75

REVIEWS

Basic Hypergeometric Series and Applications by Nathan J. Fine RANJAN ROY 82

Innumeracy: Mathematical Illiteracy and Its Consequences by John Allen Paulos LISA J. EVERED 88

TELEGRAPHIC REVIEWS 92

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN	DAVID J. HALLENBECK	LEE A. RUBEL
DONNA BEERS	PAUL R. HALMOS	ANITA E. SOLOW
DENNIS DETURCK	MELVIN HENRIKSEN	JOEL SPENCER
HAROLD G. DIAMOND	JOAN P. HUTCHINSON	LYNN A. STEEN
JOHN DIXON	WILLIAM M. KANTOR	KENNETH B. STOLARSKY
J. H. EWING	JOSEPH KONHAUSER	STAN WAGON
RICHARD GUY	RICHARD LIBERA	DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

Award for Distinguished Service to Leon Henkin

HENRY L. ALDER *University of California, Davis, CA 95616*

IVAN NIVEN *University of Oregon, Eugene, OR 97403*

The 1990 Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to Mathematics is presented to Leon Henkin for his leadership in so many different areas of mathematics education and his initiatives to increase opportunities for minorities and women to succeed in mathematics.

As a professor of mathematics at the University of California, Berkeley, Leon Henkin has an outstanding record of scholarly work in his field, logic and the foundations of mathematics. As a result, he was elected President of the Association for Symbolic Logic for a 3-year term. He served two terms as chair of his department, one of the largest in the country. Yet he has found the time and energy to make a great variety of contributions to the improvement of mathematical instruction at the precollege and college levels. He regards these activities as an integral part of his work as a university professor. He was one of the pioneers to recognize the importance of mathematicians involving themselves actively in the improvement of mathematics education at all levels. The following is only a sample of his many contributions.

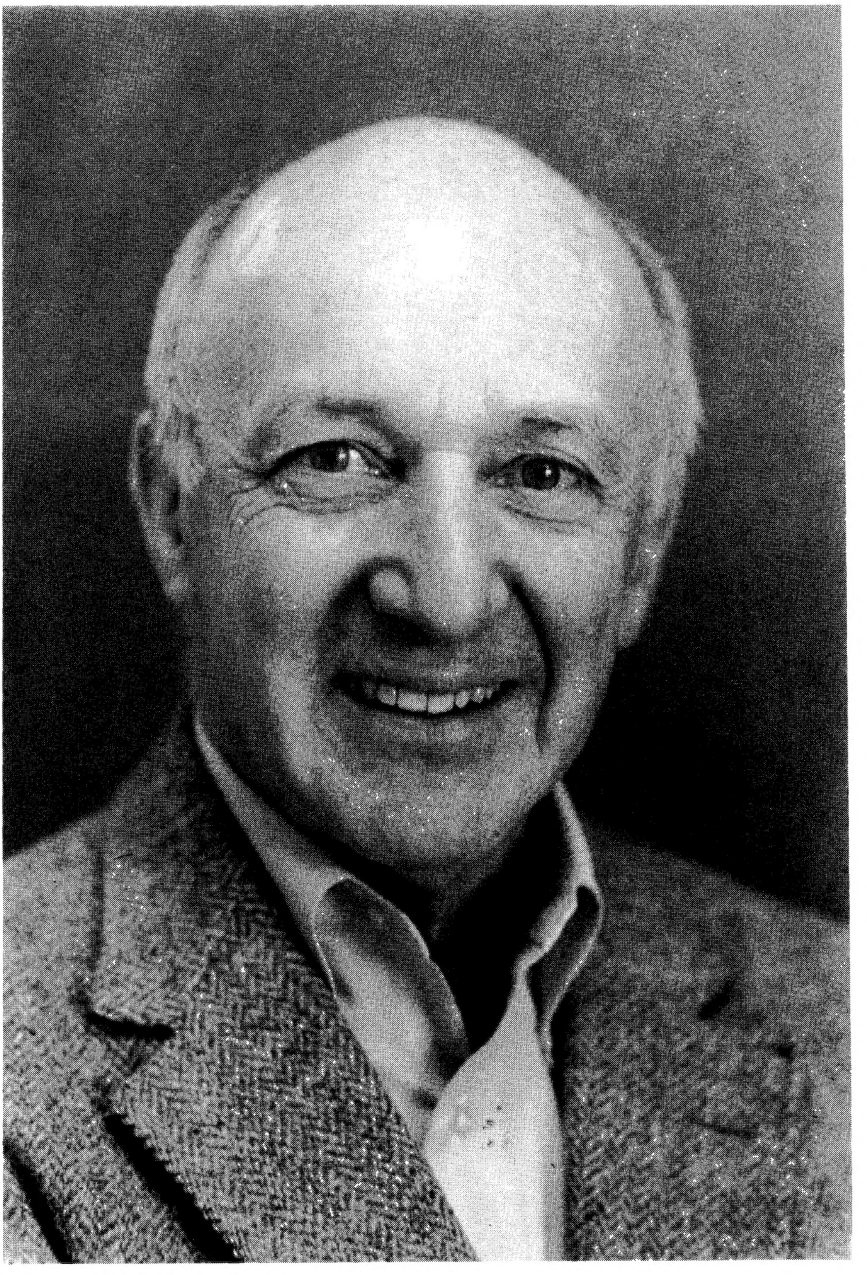
In 1964, Leon Henkin and Jerzy Neyman started a program at Berkeley to increase the number of minority students entering college from Bay Area high schools. In 1974, this was formally established as the Professional Development Program (PDP), designed to increase the number of high-achieving minority undergraduates going on to graduate or professional study leading to mathematics related careers; it is operated by a committee of the Academic Senate.

In 1969, Leon Henkin headed a statewide University of California program, still going, which sends graduate students in mathematics to teach advanced mathematical topics in elementary school classrooms on a daily basis. In 1974, Robert Davis and Leon Henkin led an NSF program to identify and develop mathematical leaders among teachers in each elementary and middle school in Berkeley.

In addition to his students in the graduate mathematics program at the University of California, Leon Henkin has had five Ph.D. students over the past twenty years in an interdisciplinary program in Mathematics and Science Education. One of these students, Uri Treisman, won the $50,000 Dana Award in Higher Education in 1987 for his successful design of PDP Mathematical Workshops, which dramatically increased the number of high-achieving calculus students from underrepresented minority groups.

Leon Henkin played a central role in the development of the Bay Area Mathematics Project (BAMP), one of the early collaborative projects linking students, teachers, parents, administrators, and community leaders to improve jointly the teaching of mathematics in the schools. Its remarkable effectiveness has made it a model for many such collaborative projects throughout the nation.

Fortunately, when BAMP started in the summer of 1983 with the first group of 50 teachers, he served on the U.S. Commission on Mathematical Instruction



LEON HENKIN

(USCMI). At that time, USCMI's almost exclusive focus was on matters pertaining to the meetings of the International Congress on Mathematical Education (ICME), held quadrennially. He urged the commission to broaden its scope by supporting a national program to create projects like BAMP in many regions. The USCMI supported his idea.

This is what led to the now-thriving American Mathematics Project (AMP), an organization designed to establish collaborative mathematics projects of the BAMP-type throughout the country. In its short existence, it has already given birth to at least 20 new projects, established a project office at the University of California, Berkeley, to assist the activities of these projects, and publishes a quarterly (*AMP-LINE*).

As a result of his leadership in establishing AMP, Leon Henkin has been called "The Father of the American Mathematics Project."

In spite of these many time-consuming activities, Leon Henkin has found time to expand his earlier initiatives to increase opportunities for minority and women students. In 1989, together with Uri Treisman, he organized a six-week summer program at the University of California, Berkeley, for such students who have completed two years of undergraduate mathematics courses beginning with calculus. This nationwide program, designed to increase the number of minority and women students seeking careers requiring a Ph.D. in mathematics, is supported by a grant from the Alfred P. Sloan Foundation. In his own department he continues to serve on the Mathematics Opportunity Committee, which has contributed significantly to the number of Ph.D. degrees granted to minority and women mathematicians since he helped to found the committee 25 years ago.

In addition, Leon Henkin has made significant contributions as a member of many key committees of the MAA and other organizations. He served with distinction on the Committee on the Undergraduate Program in Mathematics (CUPM) from 1962 to 1967, its Panel on Pre-Graduate Training from 1962 to 1965, the Committee on Educational Media from 1965 to 1967, among many others, including committees of AMS, CBMS, and NCTM.

Leon Henkin has, however, not confined his interest to projects at the national level. Recognizing the important benefits of close collaboration on all these activities also at the international level, he, together with John Kelley, took the initiative to bring the Fourth International Congress on Mathematical Education to the University of California at Berkeley in 1980. Participants from 92 countries attended the congress, the first ICME held outside Europe.

Leon Henkin is well known also for being an outstanding expositor of mathematics, both as a lecturer (in person and on film) and as a writer. He was honored with the 1964 Chauvenet Prize for outstanding expository writing for his paper "Are Logic and Mathematics Identical?," published in *Science*, 138(1962), 788–794.

Few individuals of our era have had a greater impact on the health of American mathematics than has Leon Henkin. For this, all of us owe him our deepest gratitude. Those of us fortunate enough to have been associated with him know that his service to mathematics will continue with undiminished vigor.

It is, indeed, fortunate that the first Award for Distinguished Service to Mathematics with its new title and its substantially increased financial award is given to an individual with such an exceptionally rare combination of talents that have benefited mathematics so enormously.

The Bitangent Sphere Problem

PETER J. GIBLIN, *The University of Liverpool, Liverpool L69 3BX, England*
DONAL B. O'SHEA, *Mount Holyoke College, South Hadley, MA 01075*



PETER J. GIBLIN: I did my graduate work in algebraic geometry and topology at King's College, London, and moved to the University of Liverpool in 1967. I have been there ever since, apart from years as visiting professor at Chapel Hill, N.C., and Amherst, Mass. It was in Amherst that I met Don O'Shea and somehow we couldn't let the bitangent sphere problem alone. My research interests now are in differential geometry, singularity theory, and computer vision: I am interested particularly in applying modern geometry to the problems of understanding how we perceive shape and symmetry in the world. I am also chairman of Mathematical Education on Merseyside, which, with industrial and business sponsorship, promotes competitions in mathematics for school children in the Liverpool area.



DONAL B. O'SHEA: I did my undergraduate work at Harvard and my graduate work at Queen's University at Kingston, Canada, under the supervision of A. J. Coleman. Except for a year each at I.H.E.S. and the University of Massachusetts, I have been at Mount Holyoke College since receiving my Ph.D. in 1981. My research interests are in the algebraic and differential geometry of singular complex analytic varieties, but, like Peter, I am interested in geometry, broadly defined, and its interplay with other areas of mathematics and science. I have served as chairman of the Valley Geometry Seminar and the Five College Applied Mathematics Group.

In this article, we shall introduce and solve a problem which had its origins in shape recognition. Our chief motivation for so doing is that the problem is an attractive and somewhat unusual one, and the solutions which we have found provide pleasant and instructive illustrations of techniques from differential topology. Thus, construction of differentiable manifolds (actually surfaces), vector fields, indices, "blowing up," and Morse functions will all make an appearance, some in more important roles than others.

Here is the problem: choose a point p_0 on a (smooth, closed, embedded) surface M in R^3. Does there always exist a sphere which is tangent to M at p_0 and at some *other* point $p \in M$?

Here is the answer: no. But the answer is very nearly yes, for there *is* always a *sphere or plane* having the required property; see §2.4 below for examples in which no sphere exists, but a plane does.

Here is a line of thought: start with a tiny sphere inside M, tangent to M at p_0, and pump it up until it touches M somewhere else!

Many promising arguments based on this have comé to grief. Almost irresistible is the temptation to rush on and triumplantly let $p \in M$ be the first point at which the pumped up sphere again touches M, thereby exhibiting a sphere and a point showing that the answer to the problem is yes. Other traps await those wary enough to reject the temptation, but unwilling to totally abandon the proffered line of thought. In fact, the first author made a habit for several years of asking nearly every friendly mathematician he met for a solution of the problem. Nearly all of

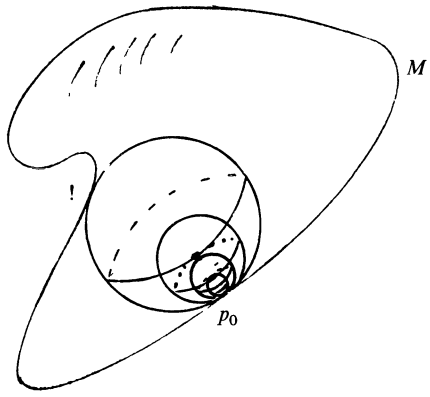


FIG. 1.

them came back rapidly with a solution, too, and the dossier of false arguments bears witness to the problem's subtlety. There is an obscure and almost certainly false argument in [3, p. 181]. We do not yet know how to extend the result to a hypersurface M^n in R^{n+1} for $n \geqslant 3$. Our attempts to extend existing arguments have run into grave technical problems (but see the endnote to §6).

The case $n = 1$ (a smooth simple closed curve in R^2) is, however, well understood and we set the scene by discussing this case in §1, where we also say a little about the origins of the problem in shape recognition.

We thank Bill Bruce, Terry Gaffney and Leo Jonker for helpful conversations. Although it would be invidious to thank here those who have contributed to the dossier of false proofs, we are nonetheless grateful to all those with whom we have discussed this problem.

1. Motivation and the curve case $n = 1$. A number of ways have been proposed for measuring or comparing the shapes of regions in the plane or in space. One method, which picks out local reflectional symmetries of the curve bounding a region in the plane (or of the surface bounding a region in space) is as follows.

For a smooth, simple closed curve M in R^2, consider circles which are *bitangent* to M, that is, tangent to M at more than one point. Such a circle picks out a "local" (first order) reflectional symmetry of the curve: the tangent lines and points of contact p_0, p_1 are symmetrical about a line (FIG. 2, left). The *centers* of all such bitangent circles trace out a locus of points called the symmetry set of M (closely related to the symmetric axis transform, medial axis, central set, ...). This set figures quite prominently in the literature of shape recognition. See [**1**, **2**] and, for more mathematical details, [**3**, **4**].

Question: can M be recovered from the bitangent circles? We need, clearly, not just the locus of their centers, but also their radii. It is certainly very plausible (FIG. 2, right) that M is just the *envelope* of the bitangent circles. But note that, if M is to be recoverable, then every point p_0 of M—or nearly every point—must contribute to the symmetry set. If a whole open arc U of M has the property that, for $p_0 \in U$, there is no bitangent circle with p_0 as a point of tangency, then there is no way of recovering U from information about bitangent circles.

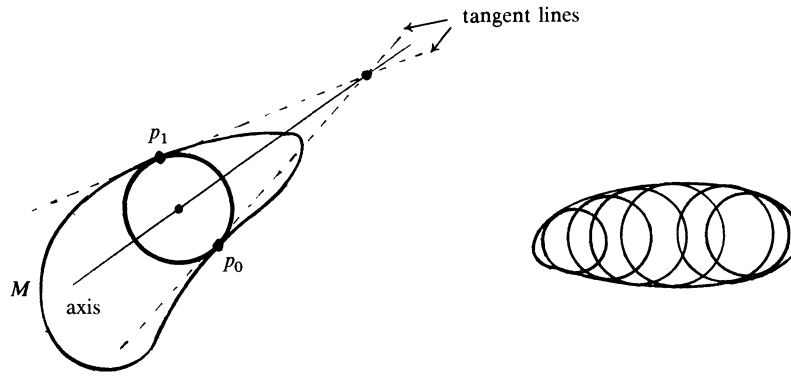


FIG. 2.

In fact, it is proved in [3, Proposition 3.5, p. 181] that, for a generic curve M (i.e., for "almost all" M), the whole of M is the *closure* of the subset M_0 reconstructed as the envelope of circles centered on the smooth part of the symmetry set: it is just necessary to fill in some limit points of M_0 to have the whole of M. This fact relies on the case $n = 1$ of our problem, namely

(1.1) PROPOSITION. *Let M be a closed curve smoothly embedded in R^2, and let $p_0 \in M$. Then there is a circle or line tangent to M at p_0 and at some point $p \in M$, $p \neq p_0$.*

Proof. Let us choose coordinates with $p_0 = (0, 0)$ and the tangent to M along the x-axis (FIG. 3). Let $\alpha : S^1 \to R^2$ be a smooth parametrization of M, $\alpha(t) = (X(t), Y(t))$, and let $\alpha(t_0) = p_0$. Note that $X(t_0) = Y(t_0) = Y'(t_0) = 0$. For $t \neq t_0$, $t \in S^1$, we write down the reciprocal of the radius $r(t)$ of the unique circle $C(t)$ through $\alpha(t)$, tangent to the x-axis at $(0, 0)$. A short calculation shows that $1/r(t)$ is

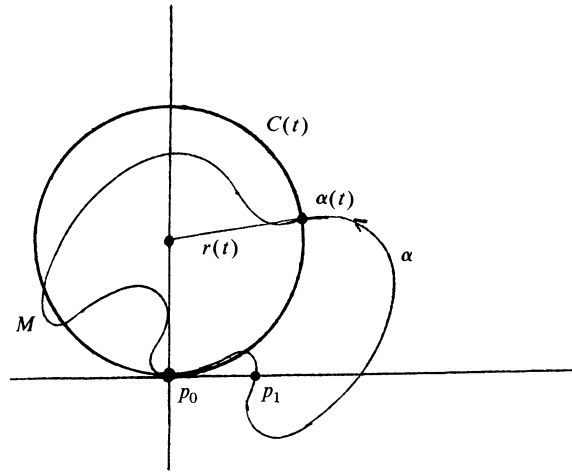


FIG. 3.

$f(t)$, where

$$f(t) = \frac{2Y(t)}{(X(t))^2 + (Y(t))^2}.$$

For a point such as $p_1 = \alpha(t_1)$ in FIG. 3, $Y(t_1) = 0$, meaning that the "circle" is a straight line. In any case, let us consider f as a function on $M - \{p_0\}$.

The condition for f to have a stationary point ($f'(t) = 0, t \neq t_0$) is, intuitively, that the radius r increase up to t and then decrease (or vice versa), which means that the circle is tangent to M at t. Indeed, it is easy to check that the normal to the above circle (or line) has direction $(2XY, Y^2 - X^2)$, and this is perpendicular to the tangent to M precisely when $f' = 0$. Hence (1.1) is equivalent to: *there exists* $t \neq t_0$ *for which* $f'(t) = 0$.

Now f can actually be extended to a smooth function on all of S^1. For two applications of l'Hôpital's rule to the quotient giving f reveal that, as $t \to t_0$, $f(t) \to Y''(t_0)/[X'(t_0)]^2$, which is, up to sign, the curvature κ of M at p_0. (Indeed, the circle $C(t)$ tends to the circle of curvature of M at t_0, as $t \to t_0$.) Thus, we define $f(t_0) = \kappa(t_0)$. By writing M locally as a graph $y = h(x) = x^2 h_1(x)$ say, for smooth h and h_1, it is then easy to check that f is a smooth function at t_0 as well as everywhere else on S^1.

Since S^1 is compact, f must achieve its maximum and minimum values, and the values of t at which this happens cannot all be at t_0. Consequently, $f'(t) = 0$ somewhere else, proving the result.

(1.2) REMARKS. (1) Note that the above proof works for any smooth α which is an *immersion* (i.e. $X'(t) = Y'(t) = 0$ never happens) so long as $\alpha(t) = p_0$ only occurs for $t = t_0$. In other words, the condition that $\alpha(S^1)$ be a *simple* curve is not used: it could have crossings except at p_0.

(2) It is of course possible for $f'(t_0)$ to be zero. The condition for this to happen is that p_0 be a *vertex* of M, that is, the curvature of M be stationary at p_0. Note that if p_0 is not a vertex of M, then there must be at least *two* values of t (different from t_0) where $f'(t) = 0$; that is, there must be at least two bitangent circles or lines which are tangent to M at p_0.

(3) Suppose p_0 is a vertex of M, and suppose f' vanishes at precisely one other point $t_1 \neq t_0$. If we also have $f(t_1) = 0$, then there is no circle, only a line, tangent to M at p_0 and elsewhere. Here is an explicit construction for a curve M where this happens (FIG. 4). Let M^* be an ellipse, and let C^* be the circle tangent to M^* at the ends p_0^* and p_1^* of the minor axis. It is easy to check that no other circle (or line) is tangent to M^* at p_0^* and elsewhere. Now *invert* the figure with center q on C^*. That is, consider the plane to be the complex plane with origin at q and consider the image of the circle and the ellipse under the map $z \to 1/\bar{z}$. Inversion carries circles passing through q into lines and all other circles into circles (see, for example, §36 of [6]). Hence, M^* becomes a non-convex curve M with bitangent line C, tangent at p_0 and p_1. Moreover, if a circle were tangent to M at p_0 and some other point, then inverting back would give a corresponding circle or line for M^*, an impossibility. Note that p_0^* is a vertex of M^*; hence, as inversion preserves contact with the circle of curvature, p_0 (also p_1) is a vertex of M.

Question. Can this procedure be modified so as to produce a whole *arc* of points on M for each of which no bitangent circle (only a line) exists? (Presumably such an M would contain two straight line segments, a highly nongeneric situation!)

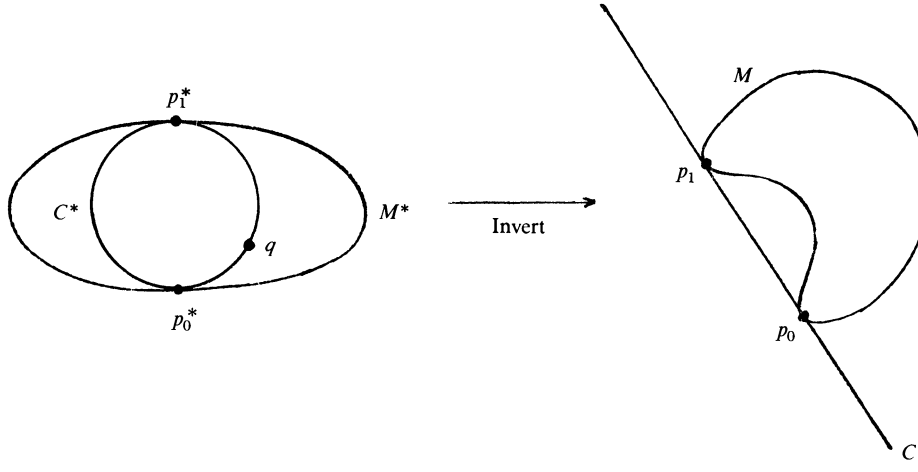


FIG. 4.

In the special case when q lies on the major axis of M^*, the curve M will have a line of symmetry. In fact, it is then a *limaçon*. Up to similarity, there is just one limaçon whose bitangent line is tangent at two vertices. In terms of its usual polar form $r = a + b \sin t$ (which can be derived from inversion of an ellipse), this is the limaçon for which $a = b\sqrt{2}$.

(4) The existence of vertices is precisely what goes wrong with the line of thought we mentioned at the outset. Given a curve M and a point p_0, any sufficiently tiny circle tangent to M at p_0 will meet M *only* at p_0. If the circle lies inside the region bounded by M it is also true that there must be circles tangent to M at p_0, obtained by pumping up the tiny circle, which meet M at points *other* than p_0. Thus, it is true that there must be a smallest circle, call it C, with the property that any smaller circle meets M only at p_0. Moreover, C will be tangent to M at any point p at which C meets M. *However*, it may happen that the only point at which C meets M is p_0 (in which case p_0 must be a vertex). An example of this is M an ellipse and p_0 one end of the major axis.

(5) We can pursue the inversion idea further to start with any M (FIG. 3) and invert with center p_0. The circles tangent to M at p_0 become horizontal straight lines, and we are looking for one of these which is tangent to M at a finite point (p_0 has disappeared to infinity). Once we know the inverted M, say M^*, has the *same* horizontal asymptote in both directions (a fact equivalent to the extendability of f in the proof of (1.1)) the result is obvious (FIG. 5).

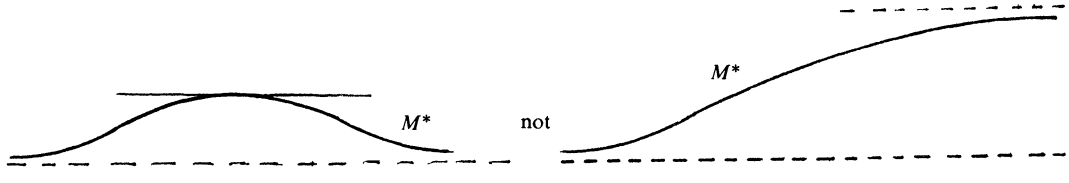


FIG. 5.

(6) The same argument used to prove (1.1) establishes the existence of a bitangent $(m-1)$-sphere or hyperplane for each point p_0 of a smooth closed curve embedded in R^m (or immersed, with no crossing at p_0).

(7) Before leaving the case $n = 1$, we mention one interesting refinement of (1.1). Some bitangent circles are undoubtedly more significant than others for the purposes of shape description. Let us orient the curve M; that is, give M a direction. In Figure 6, left, the upper circle is "coherent" in that the orientations of M at the two points of contact both point in the same way round the circle (counterclockwise), whereas the lower circle is incoherent—and probably rather insignificant, too, from the viewpoint of shape. Then (1.1) can be sharpened as follows.

(1.3) PROPOSITION. *With the hypotheses of* (1.1), *there is a coherent circle, tangent to M at p_0 and at some other point $p \neq p_0$. Indeed, if p_0 is not a vertex of M then there are two such coherent circles.*

REMARK. This result is *false* if we assume only an immersed curve M (compare Remark (1) of (1.2)). For a suitably shaped figure-eight, with p_0 at the bottom point, has only one bitangent circle and that one is incoherent (FIG. 6, middle).

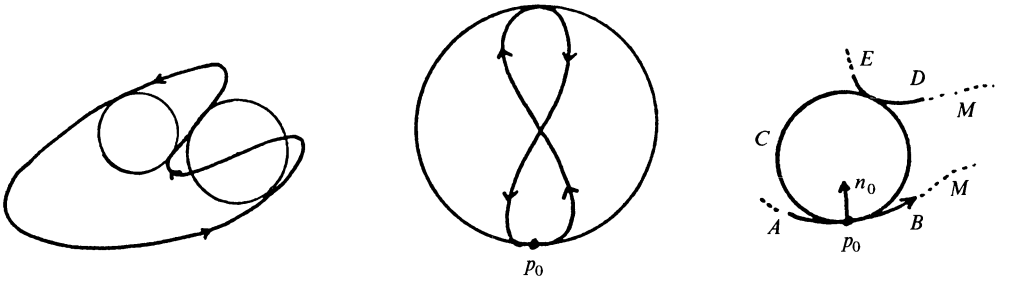


FIG. 6.

Proof of (1.3). The absolute maximum or absolute minimum of f on M, as in the proof of (1.1), will occur away from p_0, and both will do so when p_0 is not a vertex of M. Consider the absolute maximum of f, illustrated in FIG. 6, right, where n_0 is the inward-pointing normal to M at p_0. The circle C corresponding to this maximum will be entirely *inside* M. (No point on M can be inside C, for that would give a smaller circle tangent to M at p_0 and, hence, a *larger* value for f. But n_0 is the inward normal for M, so the whole interior of C is inside M.)

Let M be oriented as shown at p_0. Then we claim that the orientation of M at p is coherent (D to E in the figure). For otherwise, the inward pointing normal at p will point *out* of C, so that there will be points close to p which are *inside* C but *outside* M.

(Alternatively, it is clear that there is no way of joining B to E and A to D by arcs that do not go across the circle C.)

This proves the result for the absolute maximum, and a similar argument covers the absolute minimum case. Note that negative values of f just refer to circles which are "under" the tangent at p_0.

2. The surface case $n = 2$. The result here is:

(2.1) PROPOSITION. *Let M be a compact surface without boundary, smoothly embedded in R^3, and let $p_0 \in M$. Then there is a sphere or plane tangent to M at p_0 and at some point $p \in M$, $p \neq p_0$.*

As in the curve case, once this is known, we can deduce that a generic surface M is the closure of the subset M_0 obtained as the envelope of spheres centered on the smooth part of the symmetry set.

There is no problem setting up a function $f: M - \{p_0\} \to R$ as before. In fact, let us take $p_0 = (0,0,0)$ and the tangent plane to M at p_0 the xy-plane. Then we define $F: R^3 - \{(0,0,0)\} \to R$ by

$$F(x, y, z) = \frac{2z}{x^2 + y^2 + z^2} \tag{1}$$

and consider F restricted to $M - \{p_0\}$, which we write as f.

Let us note that *the level sets of F* are precisely the spheres centered on the z-axis and tangent to the xy-plane at $(0,0,0)$. In fact $F(x, y, z) = k$ is the sphere with center $(0,0,1/k)$ for $k \neq 0$ and the xy-plane for $k = 0$. Now *f has a critical point at p precisely when the level set of F is tangent to M at p.* (This is the principle behind Lagrange multipliers. If M is (locally) given by the equation $m(x, y, z) = 0$, then we are considering F subject to the "constraint" $m = 0$. The condition for a critical point at p is that grad $F = \lambda$ grad m at p for some Lagrange multiplier λ, and this says that the normals to the level sets $\{F = \text{const}\}$ and $\{m = \text{const}\}$ coincide at p. This is equivalent to the level sets of F and M being tangent at p.) Hence (2.1) is equivalent to:

(2.2) LEMMA. *$f = F|M + \{p_0\}$ has a critical point at some point $p \in M$, $p \neq p_0$.*

It would be nice if we could, as in the curve case, extend f to a function on M. Then we could argue that there must be a maximum *and* a minimum on M. But problems do arise when we try to extend over M. To see this, write M locally as $z = h(x, y)$ near p_0. Then f has no linear terms (the tangent plane at p_0 is $z = 0$) and we can further choose the x and y axes so that

$$h(x, y) = \frac{1}{2}\kappa_1 x^2 + \frac{1}{2}\kappa_2 y^2 + h_1(x, y) \tag{2}$$

where h_1 has order at least 3 in x and y. (The x- and y-axes are in principal directions at p_0; κ_1 and κ_2 are the principal curvatures there. This local form for M is called *Monge normal form*.)

Writing f as a function of x and y and replacing x and y by *polar coordinates* $r, \theta (x = r\cos\theta,\ y = r\sin\theta,\ r \neq 0)$ gives, after cancelling r^2,

$$f(r, \theta) = \frac{\kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta + r g_1(r, \theta)}{1 + r^2 g_2(r, \theta)} \tag{3}$$

for some smooth functions g_1 and g_2 (of period 2π in θ, of course). Now fixing θ and letting $r \to 0$, the limiting value of $f(r, \theta)$ is

$$\kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta. \tag{4}$$

If $\kappa_1 \neq \kappa_2$ then this clearly depends on θ, that is on the direction in which the origin is approached. In fact this limiting value is precisely the *sectional curvature* of M at

p_0 in the direction θ. So we cannot extend f over M unless it so happens that $\kappa_1 = \kappa_2$.

A point of M where $\kappa_1 = \kappa_2$—and hence where *all* sectional curvatures coincide with κ_1—is called an *umbilic* of M. If p_0 is an umbilic then, extending f by defining $f(p_0) = \kappa_1$, we obtain in fact a *smooth* function on all of M, and as in §1 (since M is compact) this must have a critical point away from p_0. Hence:

(2.3) LEMMA. (2.1) *holds if p_0 is an umbilic point of M.*

Question. What is the condition for f to have a critical point *at* p_0 in this case?

REMARK. Before proceeding, we remark that it is possible to have a plane, but no sphere, tangent to M at p_0 and elsewhere. In fact an example can be produced by exactly the same trick as in the curve case. Start with an ellipsoid and a sphere tangent to it at the ends of its shortest axis, then invert with center on that sphere. We hope that the reader will enjoy checking the details!

If p_0 is not an umbilic, we have seen that we cannot extend the function f on $M - \{p_0\}$ to a smooth (or even continuous) function on M. Thus, it might seem that we have reached an impasse, which is especially serious in view of the fact that umbilic points are highly nongeneric (in fact, the only closed surface on which all points are umbilic is a round sphere—see, for example, [7, Part two, V.1.2.].

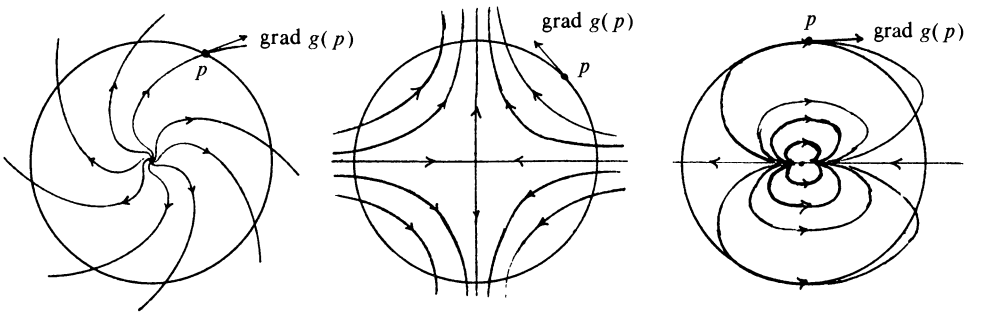
However, if we think of M as an extension of the open surface $M - \{p_0\}$, we might ask if there is some other way of extending $M - \{p_0\}$ to a closed surface which will allow us to extend f as well. Pursuing this thought will allow us to establish lemma (2.2) in the case $\kappa_1 \neq \kappa_2$.

3. The surface case: gradient vector fields. We assume from now on that $\kappa_1 \neq \kappa_2$, so that (2.3) does not apply.

The idea here is to find a compact surface $\tilde{M}$ without boundary, containing $M - \{p_0\}$ in the sense that there is a smooth injective map $i: M - \{p_0\} \to \tilde{M}$. We seek also a smooth map $\tilde{f}: \tilde{M} \to R$ extending f in the sense that $f(p) = \tilde{f}(i(p))$ for $p \in M - \{p_0\}$.

In order to prove (2.1) we shall then assume that f has *no* critical points, so that any critical points $\tilde{f}$ has must be on the "extra" part of $\tilde{M}$. By examining $\tilde{f}$ it will turn out that it has an impossible collection of critical points. So let us first see what restrictions there are on the critical points of a smooth function g on a compact surface N.

The usual way of studying the relationship between a surface (more generally, a manifold) and the critical points of a smooth function on it is through Morse theory. Morse theory is powerful and delicate, but, at least in its usual formulation, relies on the "nondegeneracy" of the critical points. (If u and v are local coordinates on N then the partial derivatives g_u and g_v are zero at a critical point of g, but $g_{uu}g_{vv} - g_{uv}^2$ is nonzero at a *nondegenerate* critical point. See [9].) There is no guarantee that our function $\tilde{f}$ has only nondegenerate critical points, so we shall work directly with the gradient vector field defined by the function. That is, at each point p of N we take the tangent vector in the direction grad $g(p)$ $(= (g_u, g_v)$ in local coordinates on $N)$. This is perpendicular to the level curve in N given by $\{g = \text{constant}\}$ through p in the direction of increasing g. For a beautiful discussion of vector fields, and of the Poincaré-Hopf theorem which we shall use, see §6 of [8]. See also [7, §3.5] and [5, §15].



index 1 (minimum) index −1 (saddle) index 2 (dipole)

The vector field at p points along the curve through p. The arrows point uphill.

FIG. 7.

Of course singular points of grad g give zeroes of the vector field. Each *isolated* zero p_1 has associated with it an *index*. To find the index, simply choose local coordinates with $p_1 = (0,0)$ and center a small circle $u^2 + v^2 = \varepsilon$ at p_1. Then see how many times the vector grad $g(p)$ turns anticlockwise as p traverses the circle once anticlockwise. This number is the index (FIG. 7). (It is also the degree of the map $p \to \text{grad } g(p)/\|\text{grad } g(p)\|$ from the circle to the unit circle.)

Suppose that N is a smooth compact surface. If N is broken up into polygonal faces meeting only along edges, then the Euler characteristic $\chi(N)$ of N is defined to be the number $V - E + F$, where V, E, and F are the number of vertices, edges, and faces, respectively. The Euler characteristic does not depend on how we break N up. Let $g: N \to R$ be a smooth function on M which has all its critical points isolated. If N has a boundary, then the vector field is required to point outwards along the whole boundary or else inwards along the whole boundary. The crucial theorem is:

(3.1) POINCARÉ-HOPF THEOREM. *The sum of the indices of the vector field is the Euler characteristic* $\chi(N)$ *of* N.

This theorem applies to any smooth vector field with isolated zeroes, not just to a gradient vector field.

In order to apply the theorem we shall need to know the index of grad g at a maximum or minimum of g. Strangely, this result (except at a *nondegenerate* maximum or minimum) does not seem to have a convenient reference, so we shall prove it.

(3.2) LEMMA. *The index of* grad g *at an isolated maximum or minimum of* g *is* 1.

Proof. Using local coordinates (u, v) we may assume that the isolated maximum or minimum is at $(0,0)$ and that $g(0,0) = 0$. It follows directly from the definition that multiplying by a nonzero constant does not change the index. Thus, upon multiplying by -1 if necessary, we may assume that $(0,0)$ is a minimum of g. Let D be a small disk centered at $(0,0)$ and containing no critical point of g besides $(0,0)$. Let δ_0 be the minimum value of g on the boundary circle S of D (note that S is compact, so this minimum is attained). We first show that if $0 < \delta < \delta_0$, then

$C = g^{-1}(\delta) \cap D$ is *one* smooth closed curve surrounding $(0, 0)$. Of course, C lies in the *interior* $\mathring{D}$ *of* D.

Now δ is certainly a regular value of g on D (the only critical point in D has value 0) so $g^{-1}(\delta) \cap D$, which is inside $\mathring{D}$, is a 1-manifold; that is, a union of smooth closed curves. Consider one of these, C_1 say. If $(0, 0)$ is not in the interior of C_1, then C_1 will bound a topological disk D_1 inside $\mathring{D}$. The function g is constant $(= \delta)$ on C_1, so g is either constant on D_1 or it achieves its absolute maximum or minimum on D_1 at a point of $\mathring{D}_1$. In either case g has a critical point inside D_1, which is not true. Similarly, if $g^{-1}(\delta) \cap D$ has two components C_1 and C_2, which we now know both contain $(0, 0)$, then the region A between C_1 and C_2 is compact and $g = \delta$ on the boundary $C_1 \cup C_2$ of A. Again g has a critical point in $\mathring{A}$, a contradiction.

Since C is a topological circle, it bounds a topological disk and a topological disk has Euler characteristic 1. Moreover, grad g points outwards all along C (indeed along the normals), so that the sole zero of grad g has index 1 by the Poincaré-Hopf theorem.

We could have avoided appealing to the Poincaré-Hopf theorem to finish the proof of the lemma above. One uses the fact that grad g is always normal to C (FIG. 8) to show that the map $C \to S^1$ (S^1 = unit circle), $p \mapsto \text{grad } g(p)/\|\text{grad } g(p)\|$, has degree 1. The degree is a homotopy invariant (this is a standard fact: see, for example, [**8**] §5) and it is easy to produce a homotopy of C onto a small circle inside C (along which grad g does not vanish). Thus, the index of the same map on a small circle inside C (and, so, the index of the zero of grad g) is 1.

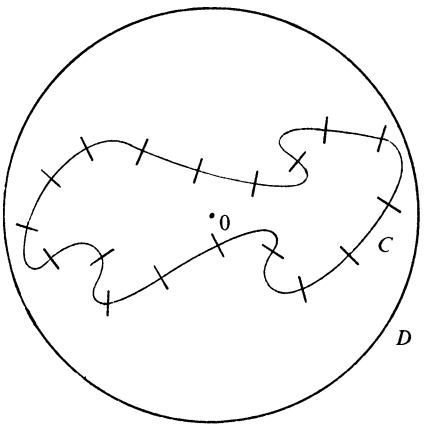


FIG. 8. Grad g points outwards from C for a minimum and inwards for a maximum.

(3.3) LEMMA. *Let* $g: N \to R$ *be a smooth function on a compact surface* N *without boundary. Then*: (i) g *cannot have exactly two maxima and two minima* (*and no other critical points*); (ii) *if* g *has exactly one maximum and one minimum* (*and no other critical points*), *then* N *is a topological sphere.*

Proof. (3.1) and (3.2) give $\chi(N) = 4$ in case (i) and $\chi(N) = 2$ in case (ii). But $\chi(N) \leqslant 2$ for any compact N without boundary, with equality if and only if N is a topological sphere (see [**10**], §39).

4. Extending M and f: doubling M. We shall give two ways of extending $f: M - \{p_0\} \to R$ to $\tilde{f}: \tilde{M} \to R$. Consider formula (3) in §2 above, which "extends f over p_0" at the price of having a different value for each direction of approach to p_0. What could be more natural than to unplug p_0 and plug in a circle S to accommodate all those different values? In fact, for diametrically opposite directions (θ and $\theta + \pi$) we do get the *same* extended value $\kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta$. How about sewing together diametrically opposite points of S, to close up the surface? That is what method 2 below does (FIG. 9, center); see §5. Alternatively, we could stick together two copies of the surface with boundary S to make a surface without boundary, by a "doubling" procedure. That is what method 1, which we shall consider in this section does (FIG. 9, right). Method 2 ("blowing up") is in a sense the least one has to do to extend M and f. Doubling is more extravagant.

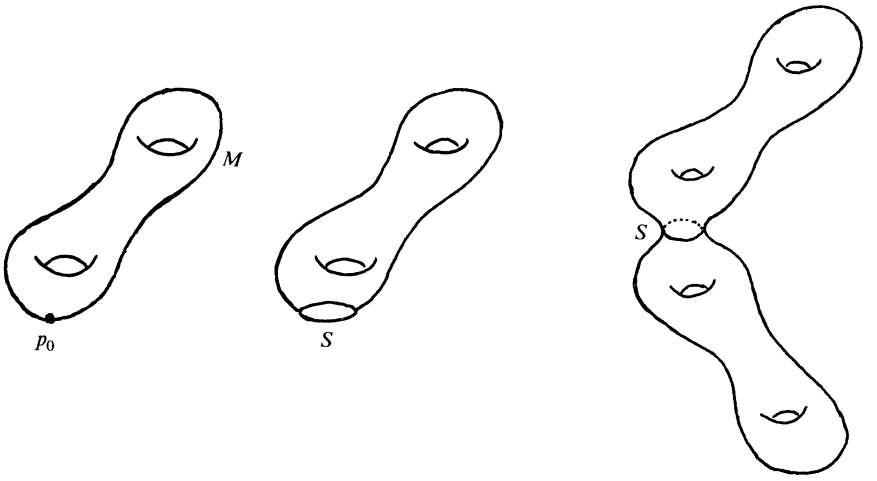


FIG. 9.

REMARK. It is tempting to work directly with the surface whose boundary is S. It turns out that f extends and that it is possible to push through a proof of lemma (2.2); in fact, this was the way that the authors first proved (2.2). However, the technical details are much more complicated. We cannot use the Poincaré-Hopf theorem because the vector field does not point all inwards (or all outwards) along S. Thus, it is necessary to appeal to Morse theory for manifolds with boundary. Even this is not straightforward: the vector field grad f has zeroes on the boundary S which is forbidden in Morse's original presentation [9] (and the subsequent formulations with which we are familiar). The arguments needed to circumvent this difficulty are delicate and neither author quite believed them until the alternate arguments detailed above were found.

Let us, as in §2, take M tangent to the xy-plane at $p_0 = (0, 0, 0)$, with M given near p_0 by $z = h(x, y)$ as in (2) of §2. In fact, let us suppose M is this graph for $0 \leqslant \sqrt{x^2 + y^2} < \varepsilon_1$. Let M_1 be the open cylinder $(-\varepsilon_1, \varepsilon_1) \times S^1$ consisting of points (r, θ) for $0 \leqslant \theta \leqslant 2\pi$ and $-\varepsilon_1 < r < \varepsilon_1$. Let M_2 and M_3 be two copies of $M - \{0\}$. We shall make a smooth surface $\tilde{M}$ out of M_1, M_2, and M_3 (FIG. 10). Now a smooth surface $\tilde{M}$ is made up of coordinate patches, that is bijective maps σ_i

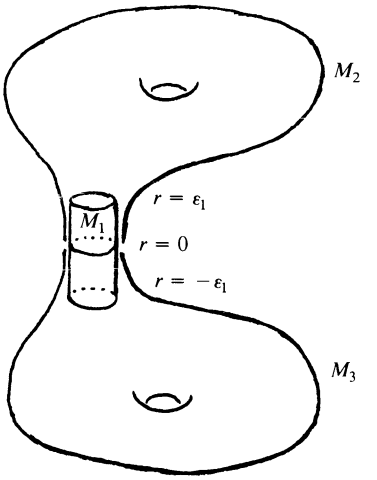


FIG. 10.

from open sets U_i in the plane to open sets $\sigma_i(U_i)$ in $\tilde{M}$. The crucial condition is that when $\sigma_i(U_i)$ and $\sigma_j(U_j)$ overlap, the "transition function" $\sigma_j^{-1} \circ \sigma_i$ must be smooth (it is a map $U \to R^2$ where U is open in R^2). When, as here, we have three known surfaces to stick together, we need

(i) to say which points (of M_1 and M_2, M_1 and M_3) are identified and

(ii) to write down transition functions, using coordinates on M_1, M_2, and M_3.

We handle these as follows:

(i) $(x, y, h(x, y)) \in M_2$ is identified with $(r, \theta) \in M_1$ where $x = r\cos\theta$, $y = r\sin\theta$ $(0 < r < \varepsilon_1)$. Also $(x, y, h(x, y)) \in M_3$ is identified with $(-r, \theta + \pi) \in M_1$.

(ii) We can use local coordinates (x, y) $(0 < x^2 + y^2 < \varepsilon_1^2)$ on M_2 and M_3. If $S^1 \subset R^2$ is $\{(X, Y): X^2 + Y^2 = 1\}$ (think of $X = \cos\theta$, $Y = \sin\theta$), then M_1 has four convenient coordinate patches:

$$\left.\begin{aligned} (X, r) &\to \left(r, X, \sqrt{1 - X^2}\right) \\ (X, r) &\to \left(r, X, -\sqrt{1 - X^2}\right) \end{aligned}\right\} \begin{matrix} |X| < 1 \\ |r| < \varepsilon_1 \end{matrix}$$

$$\left.\begin{aligned} (Y, r) &\to \left(r, \sqrt{1 - Y^2}, Y\right) \\ (Y, r) &\to \left(r, -\sqrt{1 - Y^2}, Y\right) \end{aligned}\right\} \begin{matrix} |Y| < 1 \\ |r| < \varepsilon_1 \end{matrix}$$

The transition function between M_2 and M_1 using the first coordinate patch is then the smooth function

$$(x, y) \to (X, r) = \left(\sqrt{(x^2 + y^2)}, \frac{x}{\sqrt{(x^2 + y^2)}}\right),$$

where $y > 0$ since $r = \sqrt{x^2 + y^2} > 0$ and $Y = \sqrt{1 - X^2} = \sin\theta > 0$ here. Similar

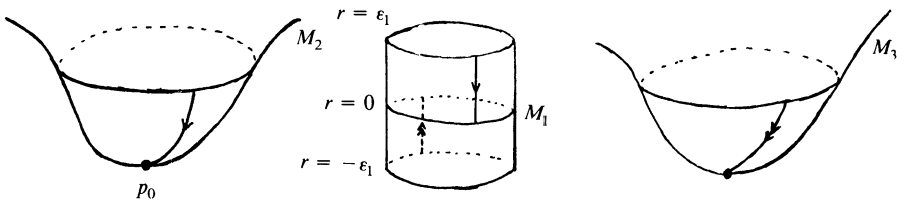


FIG. 11. Identifications to make $\tilde{M}$.

formulas apply to the other three patches. For M_3 and M_1 (first patch) we have

$$(x, y) \to (X, r) = \left(-\sqrt{(x^2 + y^2)}, \frac{-x}{\sqrt{(x^2 + y^2)}} \right),$$

where $y < 0$ since $r = -\sqrt{x^2 + y^2} < 0$ and $Y = \sqrt{1 - X^2} = \sin\theta > 0$.

In this way, we construct a smooth compact surface $\tilde{M}$ without boundary. Now we need to define $\tilde{f}\colon \tilde{M} \to R$ extending f. That is, we need to define $\tilde{f}$ on M_1, M_2 and M_3 and check the definitions agree on overlaps. The various $f_i : M_i \to R$ must also be smooth, of course.

On M_2 and M_3 we use formula (1) of §2. On M_1 we define $f_1 : M_1 \to R$ by formula (3) of §2. Thus using local coordinates (x, y) on M_2 and M_3 near p_0,

$$f_2(x, y) = f_3(x, y) = \frac{2h(x, y)}{x^2 + y^2 + (h(x, y))^2}.$$

The important property of formula (3), defining f_1, is that $f_1(r, \theta) = f_1(-r, \theta + \pi)$. This is because $r\cos\theta = (-r)\cos(\theta + \pi)$, $r\sin\theta = (-r)\sin(\theta + \pi)$ and f_1 is obtained from the same formula as f_2 or f_3 with x replaced by $r\cos\theta$, y by $r\sin\theta$ and r^2 then cancelled top and bottom. In view of this, f_1 and f_2 (resp., f_1 and f_3) do agree on the overlap of their domains, and we get a smooth function $\tilde{f}\colon \tilde{M} \to R$. It is now easy to complete the argument for (2.1).

(4.1) LEMMA. *Assume that $f : M - \{p_0\} \to R$ has no critical points. Then $\tilde{f}$ has two (absolute) maxima and two (absolute) minima, all on the $r = 0$ circle of M_1, and no other critical points.*

Proof. Certainly $\tilde{f}$ has no critical points on $\tilde{M}$ except on the $r = 0$ circle of M_1. If $\tilde{f}$ has a critical point on this circle, then, *restricting $\tilde{f}$ to the circle*, the resulting function of one variable θ will have a critical point (zero derivative). From (3) in §2, this function is

$$\theta \to \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta,$$

which (recall that $\kappa_1 \neq \kappa_2$) gives critical points precisely at $\theta = 0, \pi/2, \pi, 3\pi/2$ on the circle. Also the corresponding *values* of $\tilde{f}$ are $\kappa_1, \kappa_2, \kappa_1, \kappa_2$ respectively. Now $\tilde{f}$ must attain its absolute maximum and minimum values *somewhere* on $\tilde{M}$, and the only choice is at these four points. This proves the lemma.

Now (3.3) gives a contradiction, so (2.2), and hence (2.1), follows immediately.

Question. The idea of *coherent* bitangent circles (1) extends to coherent bitangent spheres in a natural way. Is there always such a sphere for an embedded M?

Note that the argument of (1.3) won't quite work, for possibly $\tilde{f}$ could have both its absolute maximum and minimum on the $r = 0$ circle: the other critical points might have critical values between these two. Note that then M is entirely contained between its two spheres of curvature at p_0.

5. Extending M and f: blowing up. We shall go on to give a second derivation of (2.1), because the second construction of a closed surface provides a good illustration of the important idea of "blowing up" a point. This is exactly what is needed to separate out all the different limiting values of f as $r \to 0$ (see (3), (4) in §2). Rather than separate out all *directions* (all θ) through $(0, 0, 0)$ in the plane $z = 0$, blowing up separates all *lines* (θ and $\theta + \pi$ give the same line). This is fine for us, since the limiting values of f for θ and $\theta + \pi$ ((4) in §2) coincide. The effect is (roughly speaking) to remove a neighbourhood of p_0 in M and identify the boundary of a Mobius band with the newly created boundary of M. As a result, the new surface, $\overline{M}$ say, has no boundary.

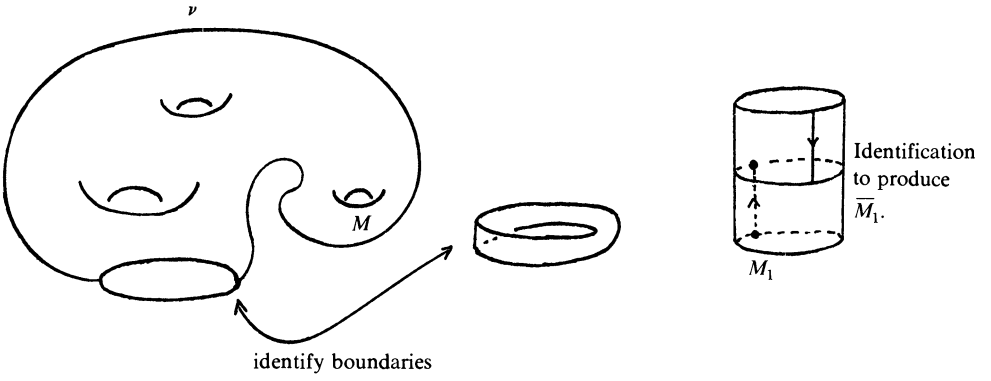


FIG. 12.

Fortunately this is very easy to do from our construction in §4 of the "double" $\tilde{M}$. We simply put together M_1 and M_2 as in §4 and then identify each pair of points (r, θ) and $(-r, \theta + \pi)$ of M_1. This has no effect on M_2, which only meets the $r > 0$ part of M_1. What we have to check is that, after these identifications, M_1 becomes a smooth manifold, $\overline{M}_1$ say, without boundary (but "open at the $r = \varepsilon_1$ end"), and that f_1 extends to $\overline{M}_1$. Note that $(0, \theta)$ and $(0, \theta + \pi)$ are identified together for each θ.

So let us concentrate on $M_1(-\varepsilon_1, \varepsilon_1) \times S^1$ and on $\overline{M}_1$, obtained by identifying (r, θ) and $(-r, -\theta)$ for each (r, θ). (Thus $\overline{M}_1$ has the "quotient topology". In particular each point has a neighbourhood homeomorphic to a 2-dimensional disk —for $r = 0$ we can picture this as two separate pieces sewn together as in FIG. 13. Thus $\overline{M}_1$ *is* a 2-manifold. Indeed $\overline{M}_1$ is topologically a Möbius band (FIG. 13, right).)

Looking back at the four coordinate patches for M_1 in §4, it is easy to see that the first pair have the *same* image in $\overline{M}_1$, and so do the second pair. We are left

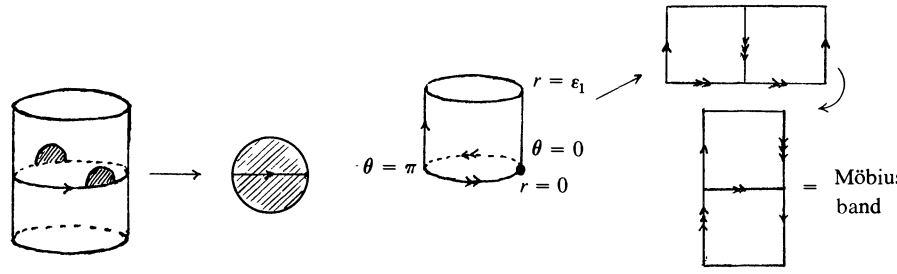


FIG. 13.

with *two* coordinate patches for $\overline{M}_1$,

$$\begin{aligned}(X, r) &\to \alpha\left(r, X, \sqrt{(1 - X^2)}\right), \qquad |X| < 1, |r| < \varepsilon_1,\\ (Y, r) &\to \alpha\left(r, \sqrt{(1 - Y^2)}, Y\right), \qquad |Y| < 1, |r| < \varepsilon_1,\end{aligned} \tag{1}$$

where α represents the natural projection map $M_1 \to \overline{M}_1$ taking $(r, \cos\theta, \sin\theta) = (r, X, Y)$ to $\{(r, X, Y), (-r, -X, -Y)\}$. These two maps are *injective* (because the square roots are positive). Furthermore given $\alpha(r, X, Y)$ in the overlap of the coordinate patches we can assume that $X > 0$ and $Y > 0$ so that the unique points going to $\alpha(r, X, Y)$ under the maps (1) are (X, r) and (Y, r). The transition function is $Y = \sqrt{(1 - X^2)}$, which is smooth on $|X| < 1$.

Thus $\overline{M}_1$ is indeed a smooth surface without boundary. We can now check that f extends to $\bar{f}_1 : \overline{M}_1 \to R$ by using the same formula ((3) in §2) as we used in §4. Thus, we get $\bar{f} : \overline{M} \to R$ which is smooth.

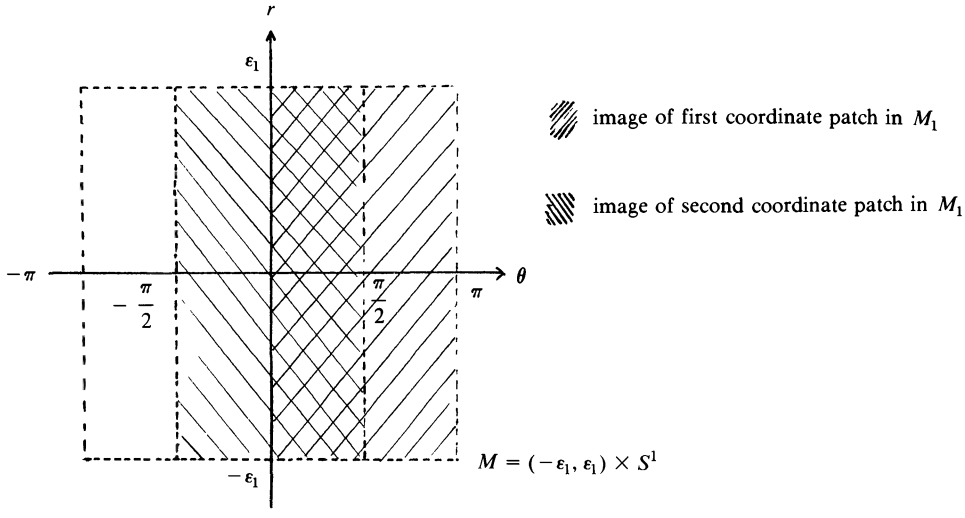


FIG. 14.

The argument is concluded in a similar way to (4.1), namely:

(5.1) LEMMA. *Assume that* $f: M - \{p_0\} \to R$ *has no critical points. Then* $\bar{f}$ *has one absolute maximum and one absolute minimum, both on the* $r = 0$ *circle of* $\overline{M}_1$, *and no other critical points.*

To deduce (2.1) we need an extra step now. For by (3.3), $\overline{M}$ is a topological sphere. But $\overline{M}_1$, and hence $\overline{M}$, contains a Möbius band (FIG. 13), which no sphere can do. This contradiction proves (2.2) and hence (2.1).

There is a more conventional way of describing the blown up part of $\overline{M}$, namely, the part coming from $\overline{M}_1$, as follows.

Let

$$U_1 = \{(x_1, y_1) \in R^2 : x_1^2(1 + y_1^2) < \varepsilon_1^2\}$$

and

$$U_2 = \{(x_2, y_2) \in R^2 : (1 + x_2^2) y_2^2 < \varepsilon_1^2\}.$$

See FIG. 15.

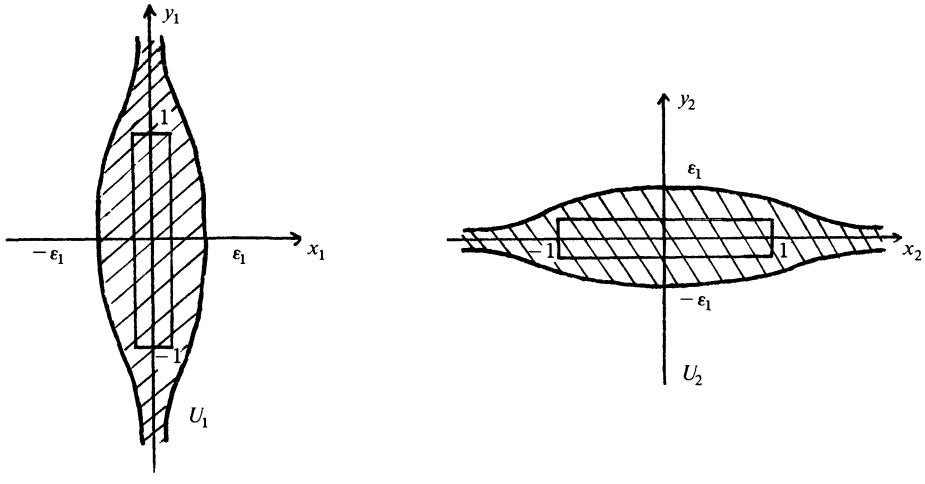


FIG. 15.

Let $\tau_1 : U_1 \to M_1$ be given by

$$\tau_1(x_1, y_1) = \left(x_1\sqrt{(1 + y_1^2)}, 1\Big/\sqrt{(1 + y_1^2)}, y_1\Big/\sqrt{(1 + y_1^2)}\right).$$

Thus $\tau_1(x_1, y_1) = (r, \cos\theta, \sin\theta)$ where $r\cos\theta = x_1$, $r\sin\theta = x_1 y_1$, and if $x_1 = 0$, the image is that part of the $r = 0$ circle for which $\cos\theta > 0$. After passing to $\overline{M}_1$ this is all of the $r = 0$ circle except for one point corresponding to $\theta = \pm\pi/2$.

Similarly let $\tau_2 : U_2 \to M_1$ be given by

$$\tau_2(x_2, y_2) = \left(y_2\sqrt{(1 + x_2^2)}, x_2\Big/\sqrt{(1 + x_2^2)}, 1\Big/\sqrt{(1 + x_2^2)}\right).$$

This is $(r, \cos\theta, \sin\theta)$ where $r\cos\theta = x_2 y_2$, $r\sin\theta = y_2$, and y_2 covers, in $\overline{M}_1$, all of the $r = 0$ circle except for one point corresponding to $\theta = 0, \pi$.

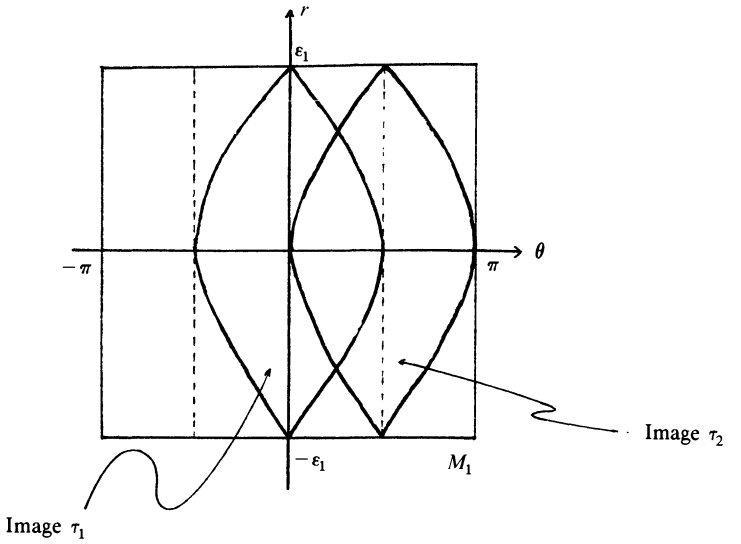


FIG. 16.

The images of τ_1 and τ_2 in M_1, and by implication in $\overline{M}_1$, are sketched in FIG. 16, which shows the cylinder M_1 opened out along the $\theta = \pm\pi$ generator. Thus $\tau_1(U_1)$ and $\tau_2(U_2)$ do not cover $\overline{M}_1$, but they do cover all of the $r = 0$ circle which is where the potential difficulties lie. There is no difficulty covering the rest of $\overline{M}_1$, for example, by covering the ordinary cylinder $\{(r, \theta): 0 < r < \varepsilon_1\} \subset \overline{M}_1$ by two coordinate patches, giving $0 < \theta < \pi$ and $-\pi/2 < \theta < \pi/2$.

The transition function between U_1 and U_2 is

$$(x_2, y_2) = (1/y_1, x_1y_1)$$

($x_1 \neq 0$, $y_1 \neq 0$ on the overlap in $\overline{M}_1$). The formula for $\bar{f}$, extending f, in terms of coordinates from U_1, is as follows. In the notation of (1) and (2) in §2, write

$$h_1(x_1, x_1y_1) = x_1^3 h_2(x_1, y_1) \text{ for some } h_2,$$

(since h_1 has order at least 3 in its coordinates). Also $\frac{1}{2}\kappa_1 x_1^2 + \frac{1}{2}\kappa_2 x_1^2 y_1^2 + (h_1(x_1, x_1y_1))^2 = x_1^4 h_3(x_1, y_1)$, for some h_3. Then

$$\bar{f}(x_1, y_1) = \frac{\kappa_1 + \kappa_2 y_1^2 + x_1 h_2(x_1, y_1)}{1 + y_1^2 + x_1^2 h_3(x_1, y_1)}.$$

Similarly for U_2, and one can check that the formulas agree on the overlap in $\overline{M}_1$. In this way, we obtain an alternative construction of $\overline{M}$ and $\bar{f}$.

Note that the images of two rectangles in FIG. 15 together make up a Möbius band in $\overline{M}_1$ and hence in $\overline{M}$.

6. What about $n \geq 3$?. We conclude with some remarks about higher dimensional cases. In the first place, there is an "easy case" analogous to (2.3). Suppose there are local coordinates $(x_1, \ldots, x_n, z) \in R^{n+1}$ centered at p_0 such that locally

M (a smooth, compact n-manifold without boundary) has the form

$$z = \frac{1}{2}\kappa_1 x_1^2 + \cdots + \frac{1}{2}\kappa_n x_n^2 + h(x_1, \ldots, x_n) \tag{1}$$

where h has order at least 3 in the coordinates and $\kappa_1 = \cdots = \kappa_n$. Then the usual function

$$F(x, \ldots, x, z) = \frac{2z}{x_1^2 + \cdots + x_n^2 + z^2},$$

restricted to $M - \{p_0\}$ can be extended over all of M. It then follows that $f = F|_M$ has a critical point away from p_0, and we deduce the existence of an n-sphere or n-plane tangent to M at p_0 and elsewhere.

This case is the "equal eigenvalues" case, referring to the quadratic form in (1). Now assume not all κ_i are equal.

We can construct smooth manifolds $\tilde{M}$ and $\overline{M}$ in a way analogous to that used in §§4 and 5, and extend f to these. In fact $\bar{f}: \overline{M} \to R$ looks the more useful one: here, p_0 is replaced by an $(n-1)$-dimensional real projective space P which separates out all lines through $p_0 = 0$ in the n-plane $z = 0$.

We assume again that f has no critical points on $M - \{p_0\}$; such a critical point would immediately give a bitangent n-sphere or n-plane, as in (2.2). Then $\bar{f}$ has critical points only on the "$r = 0$" locus P, and we can check that these can only occur at points of P corresponding to directions along one of the coordinate axes in the plane $z = 0$. The problem is that these critical points may be maxima, minima or saddles of various indices, and we run into difficulties, and many separate cases, when we try to apply the Poincaré-Hopf theorem to the vector field grad $\bar{f}$. In the case $n = 2$, we could get away without assumptions about nondegeneracy (because of (3.2) and (5.1)), but for $n \geqslant 3$ we need to be more careful.

As an example, let $n = 3$, and suppose that all the critical points of $\bar{f}$ are nondegenerate. The indices of grad $\bar{f}$ are all ± 1, and those at the maximum and minimum have opposite sign (since dim $M = 3$ is odd). Thus the sum of the three indices is ± 1, which should equal the Euler characteristic $\chi(\overline{M})$. However, $\chi(\overline{M}) = 0$ because the Euler characteristic of any 3-manifold is zero. (See, for example, Theorem III, 59 of [**10**]). In our case, this also follows directly from Poincaré duality which is applicable because $\overline{M}$ is orientable whenever M is orientable (this contrasts with the case $n = 2$, where $\overline{M}$ contained a Mobius band), and M is orientable because it is a hypersurface.

If in fact $h(x_1, \ldots, x_n) = 0$ and if all the κ_i are distinct, then we *can* show that all critical points of $\bar{f}$ are nondegenerate, but that is rather a special case!

Endnote. At a late stage in the preparation of this article, the first author discussed the problem with Elmer Rees of the University of Edinburgh. Prof. Rees and a visitor from China, Duan Hai Bao, have subsequently announced a proof in the case $n \geqslant 3$ without nondegeneracy assumptions. The argument they give (in "The Existence of Bitangent Spheres", Proc. Royal Soc. Edinburgh 111A (1989), 85–87) is based on much more technical considerations, related to the Lusternik-Schnirelmann category of a manifold, than those discussed in this paper. We still do not know of an "elementary" argument which works in the general case.

REFERENCES

1. H. Blum, Biological shape and visual science I. *J. Theoret. Biol.* 38 (1973) 205–287.
2. M. Brady, Criteria for representations of shape, in 'Human and Macine Vision', ed. by J. Beck et al., Academic Press, New York, 1983.
3. J. W. Bruce, P. J. Giblin, and C. G. Gibson, Symmetry sets, *Proc. Royal Soc. Edinburgh*, 101A (1985) 163–186.
4. P. J. Giblin and S. A. Brassett, Local symmetry of plane curves, *Amer. Math. Monthly*, 92 (1985) 689–707.
5. V. Guillemin and A. Pollack, Differential Topology, Prentice-Hall, Englewood Cliffs, NJ, 1974.
6. D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, translated by P. Nemenyi, Chelsea, New York, 1983.
7. H. Hopf, Differential Geometry in the Large, Lecture Notes in Mathematics, 1000, Springer-Verlag, New York, 1983.
8. J. W. Milnor, Topology from the Differentiable Viewpoint, Univ. of Virginia Press, 1965.
9. M. Morse and G. B. van Schaack, The critical theory under general boundary conditions, *Ann. Math.*, 35, no. 3 (1934) 545–571.
10. H. Seifert and W. Threlfall, A Textbook of Topology, translated by M. Goldman, Academic Press, New York, 1980.

The σ-Game and Cellular Automata

KLAUS SUTNER, *Stevens Institute of Technology, Hoboken, NJ 07030*

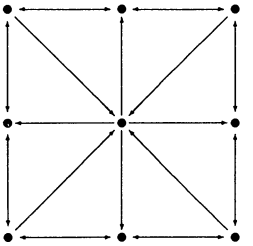
KLAUS SUTNER received his doctorate in Mathematics from the Ludwig Maximilians Universität in Munich in 1984. He is currently teaching Computer Science at Stevens Institute of Technology in Hoboken, New Jersey. His research interests lie in the area of the theory of computation.



1. Summary. In an article in this journal Don Pelletier discussed the mathematics involved in a little battery operated toy called MERLIN (see [3], and also the "Addenda" in this MONTHLY, Dec. 1987, page 994). Several years ago Stephen Wolfram, in another article that appeared in the MONTHLY, analyzed a number of simple cellular automata and the fractal patterns generated by some of these automata (see [6]). In this article we point out the close connection between MERLIN-type games and a class of cellular automata related to the ones described by Wolfram. We introduce a game played on directed graphs and give a detailed analysis of the special case where the graph is a rectangular grid. Our analysis uses linear algebra as well as ideas from the theory of cellular automata.

2. Merlin's Magic Square and σ-automata. We begin with a description of a simple combinatorial game called the σ-game. The σ-game is played on a directed graph G. We suppose that the vertices of G can have one of two different states, which we designate by 0 and 1. A configuration is an assignment of states to all the vertices. A move in the game consists of the player picking a vertex. One may think of a vertex v in G as a button that the player can press at his discretion. If vertex v is chosen the states of all the vertices u such that (v, u) is an edge in G will change. Now suppose the opponent picks two configurations on G, say X_s (the source configuration and X_t (the target configuration). To win the player has to find a sequence of moves that transforms configuration X_s into configuration X_t. Any such sequence will be called a *solution* for (X_s, X_t). Call (X_s, X_t) a *winning pair* iff there is a solution for (X_s, X_t). Similarly one defines the σ^+-game: the move v changes the state of vertex v itself in addition to changing the states of all adjacent cells.

A battery operated toy called MERLIN that emulates the σ^+-game on the nine point directed graph G_M shown below is described in [3]. X_t is fixed and X_s is randomly generated by the device.



In another version of the game the target configuration X_t is fixed to be the "All-Ones" configuration **1** and the source configuration X_s is always **0**. However, the underlying graph varies, say, over all $n \times n$ grids. In order to extend the definition of the σ-game to undirected graphs such as grids, let us adopt the convention that every undirected edge $\{u, v\}$ is interpreted as the pair of directed edges (u, v) and (v, u).

The following questions come to mind:

- Given the underlying graph G, what are the winning pairs?
- Given a winning pair (X_s, X_t), how many solutions are there; in particular, when is the solution unique?
- Given a winning pair (X_s, X_t), what is a minimal solution, i.e., a shortest sequence of moves that will transform X_s into X_t?

The reader may have noticed that the σ-game can be described using linear algebra on $\mathbf{F}_2$, the field with two elements. Let $G = \langle V, E \rangle$ be a finite directed graph on n points. One may think of the adjacency matrix A of G as an n by n matrix over $\mathbf{F}_2$: the matrix A has a 1 in the u-th row and vth column iff there is a directed edge from vertex u to vertex v in G. A configuration X is a map from V to $\mathbf{F}_2$. The collection of all configurations of G is the *configuration space* of G, in symbols C_G. Configurations are conveniently identified with subsets of V, namely X is identified with $\{v \in V | X(v) = 1\}$. C_G may be construed as a vector space over $\mathbf{F}_2$; addition here amounts to taking symmetric differences. We will write **0** for the empty set and **1** for the set V as members of C_G, so that $\mathbf{0}(v) = 0$ and $\mathbf{1}(v) = 1$ for all v in V.

Now define a map $\mu\colon V \times C_G \to C_G$ by

$$\mu(v, X)(u) := \begin{cases} 1 + X(u) & \text{if } (v, u) \text{ is an edge in } G, \\ X(u) & \text{otherwise.} \end{cases}$$

Thus $\mu(v, X)$ is the configuration that results from configuration X after move v. μ is readily extended to sequences of moves. It is straightforward to see that

$$\mu(vv, X) = X$$

and

$$\mu(uv, X) = \mu(vu, X).$$

Therefore, a solution to the game is a subset of V rather than a sequence in V. In other words, μ may be construed as a map $\mu\colon C_G \times C_G \to C_G$. Let us introduce a map

$$\sigma\colon C_G \to C_G$$

defined by

$$\sigma(Z) := \mu(Z, \mathbf{0}).$$

Note that $\mu(v, X) = X + \sigma(v)$. Hence $\mu(Z, X) = X + \sigma(Z)$. A solution to the σ-game is thus given by a configuration Z such that $\mu(Z, X_s) = X_t$, or, equivalently,

$$\sigma(Z) = X_s + X_t. \tag{2.1}$$

Observe that by (2.1) (X_s, X_t) is a winning pair iff (X_t, X_s) is a winning pair, a fact that is somewhat less than obvious from the original definition of the σ-game. One

can define μ^+ and σ^+ similarly. If the underlying graph is not clear from context we will occasionally write σ_G and so forth.

The key observation that makes it possible to analyze our game in great detail is that both σ and σ^+ are linear. To see this think of the adjacency matrix A of G as a n by n matrix over $\mathbf{F}_2$ as described above. Now construe a configuration X as a column vector over $\mathbf{F}_2$. Then clearly

$$\sigma(X) = A^T \cdot X \quad \text{and} \quad \sigma^+(X) = (A^T + I) \cdot X, \tag{2.2}$$

where A^T is the transpose of the adjacency matrix of G and I is the identity matrix. Hence σ and σ^+ are linear maps from the configuration space to itself. Thus we have the following preliminary answer to the first two questions above: (X_s, X_t) is a winning pair iff X_t lies in the affine subspace $X_s + rg(\sigma)$ where $rg(\sigma)$ denotes the range of σ. Furthermore, the collection of all solutions is an affine subspace of C_G and has the form $Z_0 + \ker(\sigma)$ where Z_0 is any particular solution for (X_t, X_s) and $\ker(\sigma)$ is the kernel of σ. Let us define $d(G) := \dim(\ker(\sigma)) = \log_2(|\ker(\sigma)|)$. Similarly let $d^+(G) := \dim(\ker(\sigma^+))$. Then any winning pair of the σ-game on G has exactly $2^{d(G)}$ solutions and there are $2^{d^+(G)}$ solutions for the σ^+-game.

To find a solution of the σ-game therefore amounts to solving a system of linear equations over $\mathbf{F}_2$, and in order to determine the number of solutions one has to compute the corank of the adjacency matrix of G. We will give a simple example of this procedure. Let us agree on some notation for graphs: P_m will denote the undirected path graph on m points. The vertices of P_m are assumed to be $\{1, 2, \ldots, m\}$. For later use define $P_{m,n}$ to be the rectangular grid on vertices $\{1, \ldots, m\} \times \{1, \ldots, n\}$. Note that the adjacency matrix A of P_m has 1's in the sub- and the super-diagonal, and 0's everywhere else. A little computation shows that:

$$A \text{ has rank } m \text{ iff } m \text{ is even}$$

and

$$A \text{ has rank } m - 1 \text{ iff } m \text{ is odd.}$$

Thus for even m the range of σ is C_G by (2.2). It follows from (2.1) that every pair (X_s, X_t) is winning on P_m for even m. For odd m one can see that the range of σ consists of all configurations Y such that $|Y \cap \{1, 3, \ldots, m\}|$ is even. Hence, for m odd, (X_s, X_t) is winning iff $|(X_s + X_t) \cap \{1, 3, \ldots, m\}|$ is even. Note that $\{1, 3, \ldots, m\}$ is the only nontrivial element in the kernel of σ in this case. Therefore, we have shown

$$d(P_m) = \begin{cases} 0 & \text{if } m \text{ is even,} \\ 1 & \text{if } m \text{ is odd.} \end{cases} \tag{2.3}$$

Similarly one can show that

$$d^+(P_m) = \begin{cases} 0 & \text{if } m \not\equiv 2 \pmod 3, \\ 1 & \text{if } m \equiv 2 \pmod 3. \end{cases} \tag{2.4}$$

Unfortunately this method breaks down when applied to general grids: the adjacency matrix of $P_{m,n}$ becomes unwieldy and there seems to be no elegant way of solving the equations of the form $A \cdot X = Y$. Of course, one may use brute force and a computer.

However, there is a different approach to the σ-game in terms of cellular automata. A cellular automaton is a discrete dynamical system. Its components,

called cells, are interconnected in a fixed way and can assume finitely many possible states. Here we will be interested only in automata with possible states 0 and 1. A configuration of the automaton is a map from its cells to $\{0, 1\}$. Every configuration determines a next configuration via a local transition rule: the state of a cell at time $t + 1$ depends only on the states of its neighbors at time t. Now suppose the underlying mesh of cells of an automaton is given by a directed graph $G = \langle V, E \rangle$. Define the *neighborhood of* v in G to be

$$\Gamma(v) := \{u \in V | (u, v) \in E\}.$$

Note that this is the open neighborhood of v, i.e., v is not included in $\Gamma(v)$. We write $\Gamma^+(v)$ for the closed neighborhood $\Gamma(v) \cup \{v\}$. Let X be a configuration on G, $X: V \to \mathbf{F}_2$. Then we may express the map σ by:

$$\sigma(X)(v) = \sum_{u \in \Gamma(v)} X(u).$$

Hence σ can be construed as a cellular automaton rule. The dynamical system consisting of the graph G together with the rule σ will be called the *σ-automaton* on G. Similarly one obtains rule σ^+ with $\Gamma^+(v)$ taking the place of $\Gamma(v)$. Observe that the rules σ and σ^+ are well defined even for infinite graphs as long as the graphs are locally finite (i.e., $\Gamma(v)$ is finite for all v).

A well-known example of a σ-automaton is the traditional 1-dimensional cellular automaton presented in Wolfram's article. The underlying (undirected) graph is the infinite path P_∞. Thus there are infinitely many cells v_i, $i \in \mathbf{Z}$, and the neighborhood of cell v_i consists of v_{i-1} and v_{i+1}. Rules σ and σ^+ on P_∞ are usually called rule 90 and rule 150 respectively. The evolution of the seed configuration $X := \{v_0\}$ under rule σ and its relation of binomial coefficients are discussed by Steven Wolfram in [**6**], see also [**7**] and [**8**]. The resulting fractal pattern for rule σ has dimension $\log_2 3$ and for rule σ^+ $\log_2(1 + \sqrt{5})$.

The existence of predecessor configurations on finite σ-automata is connected with the reversibility of the automaton. The rule σ is said to be *reversible* on G if every configuration X in C_G can be reconstructed from $\sigma(X)$, in other words, if σ_G: $C_G \to C_G$ is injective. Because σ_G is a linear map this is equivalent with σ_G having trivial kernel. But C_G is a finite space, thus rule σ is reversible on G iff σ_G is surjective. Hence every pair of configurations is a winning pair on G iff every winning pair on G has exactly one solution. For the Merlin graph G_M from above it is shown in [**3**] that rule σ^+ is reversible. Thus every configuration on G_M has a predecessor and the game MERLIN always has a uniquely determined solution regardless of which starting configuration the device generates.

A configuration on a cellular automaton that fails to have a predecessor is often called a Garden-of-Eden configuration: once lost it cannot be regained. See [**1**] and [**2**] for a discussion of Garden-of-Eden configurations in classical cellular automata. If the underlying graph G is undirected, one can show that a configuration X has a predecessor under rule σ iff

for every configuration Y in the kernel of σ, the cardinality of $X \cap Y$ is even. (2.5)

Incidentally, the same result holds for rule σ^+. As a consequence, σ^+-automata on undirected graphs have the peculiar property that the All-Ones configuration $\mathbf{1}$ is never a Garden-of-Eden configuration. Hence $(\mathbf{0}, \mathbf{1})$ always is a winning pair for the

σ^+-game on undirected graphs. Using standard techniques from linear algebra one can solve the system $A \cdot X = \mathbf{1}$ and compute a basis for the affine subspace of all solutions in polynomial time (i.e., in a number of steps polynomial in the number of vertices of G). However, it is NP-hard to find a minimal solution. Hence most likely the only way to construct a minimal solution is to perform an exhaustive search over all $2^{d^+(G)}$ candidates. For a proof see [**4**] and [**5**]. To determine the number of solutions is similarly quite difficult. Even for a simple class of graphs such as rectangular grids we do not know how many solutions there are in general. Figure 1 gives solutions for the 48×48 and the 80×80 square. The solutions are unique in each case.

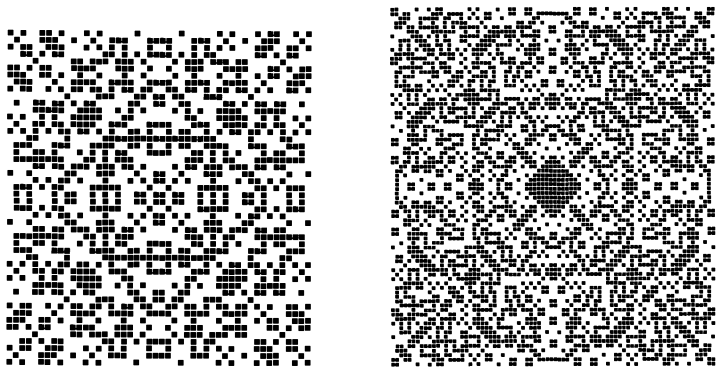


FIG. 1. Solutions to the all-ones problem for 48×48 and 80×80 squares (using rule σ^+).

It is easy to see that in general $(\mathbf{0}, \mathbf{1})$ fails to be a winning pair for rule σ. For example on the 3×3 square $P_{3,3}$ configuration $\mathbf{1}$ cannot be reached from $\mathbf{0}$ by any sequence of moves. Indeed, $Y := \{(1,1),(2,2),(3,3)\}$ lies in the kernel of σ on $P_{3,3}$. But $|Y \cap \mathbf{1}|$ is odd, and hence by (2.5), $\mathbf{1}$ fails to have a predecessor.

For rule σ, however, it is relatively easy to determine the number of solutions on a m by n grid. The remainder of this note is devoted to a proof of the following theorem.

THEOREM. *For all $m, n \geqslant 1$ the kernel of rule σ on the $m \times n$ grid $P_{m,n}$ has dimension*

$$\gcd(m+1, n+1) - 1.$$

Hence the $m \times n$ grid is reversible under rule σ iff $m+1$ and $n+1$ are relatively prime.

In fact we will explicitly construct a basis for the kernel of σ on $P_{m,n}$. The results in [**4**] show that this description of the kernel of σ on $P_{m,n}$, in connection with (2.5), implies that $(\mathbf{0}, \mathbf{1})$ is a winning pair on the $n \times n$ square iff n is even. The number of solutions in this case is 2^n by the theorem. In general let $p := \gcd(m+1, n+1) - 1$. Then the All-ones game has a solution on $P_{m,n}$ with rule σ iff

(1) p is even or

(2) $(m+1)(n+1)/(p+1)^2$ is even.

3. The number of solutions on rectangular grids. To prove the theorem we will exploit the fact that it is frequently possible to simulate one cellular automaton on another. The evolution of a configuration on the first automaton can thus be studied on the second automaton. To make this precise, let us say that the σ-automaton on H *simulates* the σ-automaton on G iff there is an injective linear map $f\colon C_G \to C_H$ such that for all X in C_G:

$$f(\sigma_G(X)) = \sigma_H(f(X)).$$

The map f will be called a simulation of G on H. Note that since f is required to be injective we have $d(G) \leqslant d(H)$ whenever H simulates G. As an example consider the path on m points P_m. The σ-automaton on P_{2m+1} simulates the automaton on P_m via the map $f(x) := \{x, 2m + 2 - x\}$. Note that for any configuration X on P_m and all $t \geqslant 0$ $\sigma^t(f(X))(m + 1) = 0$, by the symmetry of rule σ. More generally, we have the following proposition. Define two involutions (reflections on the first and second coordinates) of $P_{m,n}$ by $F_1(x, y) := (m - x + 1, y)$ and $F_2(x, y) := (x, n - y + 1)$. For any configuration X on $P_{m,n}$ let X_j denote the j-th "column" of X: $X_j := X \cap \{1, \ldots, m\} \times \{j\}$, $1 \leqslant j \leqslant n$.

PROPOSITION 1. *The σ-automaton on $P_{m,n}$ can be simulated by $P_{m,2n+1}$. Hence $d(P_{m,n}) \leqslant d(P_{m,2n+1})$ and the σ-automaton on $P_{m,2n+1}$ is reversible only if the σ-automaton on $P_{m,n}$ is reversible. If in addition P_m is reversible then $P_{m,2n+1}$ is reversible iff $P_{m,n}$ is reversible. The same holds for rule σ^+.*

Proof. The map $f\colon C_{P_{m,n}} \to C_{P_{m,2n+1}}$ defined by

$$f(X)(x, y) := \begin{cases} 0 & \text{if } y = n + 1 \\ X(x, y') & \text{if } y' = y \text{ or } y' = 2n + 2 - y,\ 1 \leqslant y' \leqslant n. \end{cases}$$

is a simulation. For example, $\sigma(f(X))(x, n + 1) = f(X)(x - 1, n + 1) + f(X)(x + 1, n + 1) + f(X)(x, n) + f(X)(x, n + 2) = X(x, n) + X(x, n) = 0$. Hence the kernel of σ on $P_{m,2n+1}$ contains the image of the kernel of σ on $P_{m,n}$ under f. Clearly $f(X) \neq \mathbf{0}$ whenever $X \neq \mathbf{0}$ and the first claim follows.

Now suppose P_m is reversible under rule σ. We will show the image of the kernel of $\sigma_{P_{m,n}}$ under f consists precisely of those configurations in the kernel of

$$\sigma_{P_{m,2n+1}}$$

that are invariant under $\mathbf{F}_2$. Pick such a configuration Y in the kernel of

$$\sigma_{P_{m,2n+1}}$$

that is invariant under $\mathbf{F}_2$. Then for all x, $1 \leqslant x \leqslant m$, $Y(x, n) = Y(x, n + 2)$. Hence $\sigma_{P_m}(Y_{n+1}) = \mathbf{0}$ and $Y_{n+1} = \mathbf{0}$ by our assumption that σ is reversible on P_m. Set

$$X := Y_1 \cup \cdots \cup Y_n \subseteq P_{m,n}.$$

Then X is in the kernel of σ on $P_{m,n}$ and clearly $f(X) = Y$. It remains to show that whenever the kernel of σ on $P_{m,2n+1}$ is nontrivial then there exists a non-trivial configuration in the kernel that is invariant under $\mathbf{F}_2$. Let $Y \neq \mathbf{0}$ be in the kernel of σ. If Y is invariant under $\mathbf{F}_2$ we are done, otherwise set $Y' := Y + \mathbf{F}_2(Y)$. Then Y' is nontrivial, lies in the kernel of σ and is a fixed point under $\mathbf{F}_2$. The argument for rule σ^+ is entirely similar and will be omitted. ∎

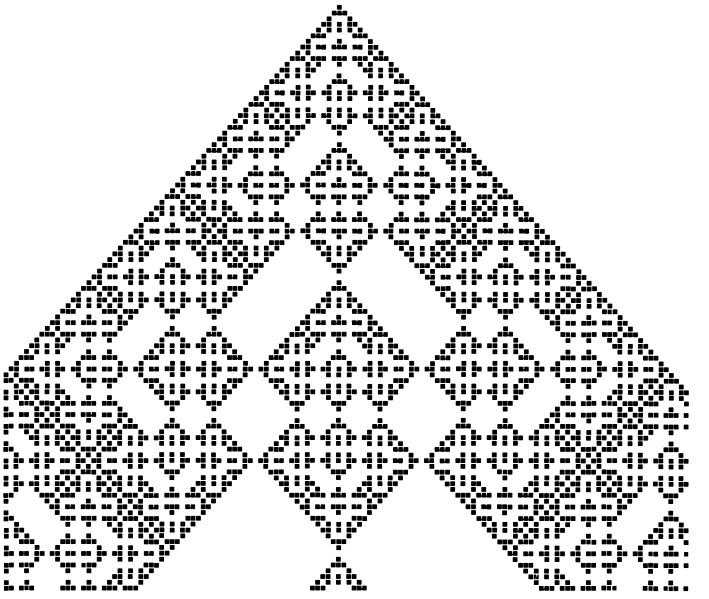


FIG. 2. Extending initial configuration $X_1 = \{0\}$ on P_∞ with rule σ^+.

It follows from proposition 1 by induction on k that every $(2^k - 1) \times (2^k - 1)$ square grid is reversible under rule σ^+: the path on $2^k - 1$ points is reversible (since $2^k - 1 \not\equiv 2 \pmod 3$). In fact any r-dimensional hypercube of the form

$$(2^{k_1} - 1) \times (2^{k_2} - 1) \times \cdots \times (2^{k_r} - 1)$$

is reversible under rule σ^+. The σ^+-game on any such hypercube therefore has only winning-pairs and the solution is always unique.

We will now give a useful description of the kernel of σ on the grids $P_{m,n}$. First note that a configuration X in the kernel of σ is completely determined by its first column X_1. To lighten notation, we write just τ for the rule σ on the path graph P_m. Then:

$$\begin{aligned} X_1 &= \mathrm{id}(X_1) = \tau^0(X_1), \\ X_2 &= \tau(X_1) = \tau^1(X_1), \\ X_3 &= \tau(X_2) + X_1 = \tau^2(X_1) + \tau^0(X_1), \\ X_4 &= \tau(X_3) + X_2 = \tau^3(X_1) \end{aligned}$$

and so forth. More precisely the sequence $(X_j\colon\ j \geqslant 1)$ of configurations is defined inductively by

$$\begin{aligned} X_1 &= \mathrm{id}(X_1), \\ X_2 &= \tau(X_1), \\ X_j &= \tau(X_{j-1}) + X_{j-2} \quad \text{for } j \geqslant 3. \end{aligned} \tag{3.1}$$

For example letting $m = 5$ and $X_1 = \{2,3,5\}$ we get successively: $X_2 = \{1,2,3\}$, $X_3 = \{1,2,4,5\}$, $X_4 = \{3,4,5\}$, $X_5 = \{1,3,4\}$ and $X_6 = \mathbf{0}$. X_n can clearly be

written as a polynomial in τ applied to X_1:

$$X_n = \sum_{0 \leqslant j \leqslant n} c_{n,j} \cdot \tau^j(X_1).$$

The two-dimensional pattern of the coefficients $c_{n,j}$ displays self-similarity properties. In fact, it has fractal dimension $\log_2 3$ and resembles the pattern obtained from a one-point seed on P_∞ and rule σ after a rotation. A picture of the latter can be found in Wolfram's article.

Now suppose we start with an arbitrary partial configuration X_1 in the first column. Define X_2, X_3 and so forth as above and let

$$X := X_1 \times \{1\} \cup X_2 \times \{2\} \cup \cdots \cup X_n \times \{n\} \subseteq P_{m,n}.$$

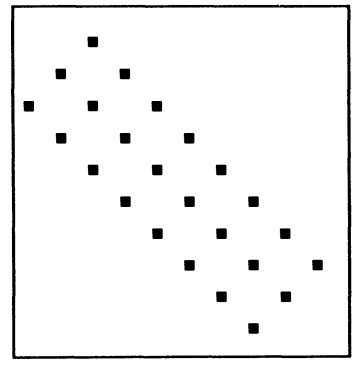
It is obvious from the definition of X_j that for all $1 \leqslant i \leqslant m$, $1 \leqslant j < n$: $\sigma(X)(i,j) = 0$. Thus X is in the kernel of σ iff $\tau(X_n) + X_{n-1} = \mathbf{0}$. Hence we have the following proposition.

PROPOSITION 2. *A configuration X' in C_{P_m} can be extended to a configuration X in $C_{P_{m,n}}$ such that $X_1 = X'$ and $\sigma(X) = \mathbf{0}$ in at most one way. The only possible extension is described by* (3.1) *and is successful iff $X_{n+1} = \mathbf{0}$.*

It is easy to see that X_{n+1} is a linear function of $X_1 = X'$, hence $d(P_{m,n})$ is the corank of the map $X_1 \mapsto X_{n+1}$. Therefore $d(m, n) \leqslant n$ and by symmetry

$$d(P_{m,n}) \leqslant \min(m, n). \tag{3.2}$$

Let us define the following "checkerboard" configurations $Z_{k,m}$ on $P_{m,m}$. $Z_{k,m}(k + i - j, i + j + 1) := 1$ for $i = 0, \ldots, m - k$, $j = 0, \ldots, k - 1$, and $Z_{k,m}(x, y) := 0$ everywhere else. The following picture shows $Z_{3,10}$.



We claim that $\{Z_{k,m} | 1 \leqslant k \leqslant m\}$ is a basis of the kernel of σ on $P_{m,m}$. It is easy to see that $\sigma(Z_{k,m}) = \mathbf{0}$. The configurations $Z_{k,m}$ are linearly independent in C_G. But by (3.2) $d(P_{m,m}) \leqslant m$ and we are done.

Using a simulation, we can now establish a lower bound on $d(P_{m,n})$. To this end let $p + 1$ be a divisor of $m + 1$ and $n + 1$. We will simulate the $p \times p$ square on $P_{m,n}$. Note that there is a tiling of $P_{m,n}$ by $p \times p$ squares that leaves gaps of width 1 between the squares as shown below. The simulation f: $P_{p,p} \to P_{m,n}$ will place one copy of configuration X into each of the squares in the tiling. To insure that f is a

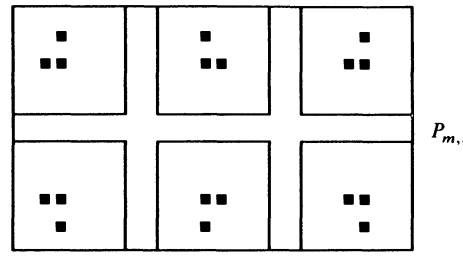
simulation some of the copies will be horizontal or vertical reflections of X. More precisely, the (i, j)th square contains $\mathbf{F}_2^{i+1 \pmod 2}(F_1^{j+1 \pmod 2}(X))$. Formally define f as follows:

$$f(a, b) := \{(a_i + i \cdot (p+1), b_j + j \cdot (p+1)) \mid 0 \leqslant i < m_0, 0 \leqslant j < n_0\} \quad (3.3)$$

where $m_0 := (m+1)/(p+1)$, $n_0 := (m+1)/(p+1)$ and

$$a_i := \begin{cases} a & i \text{ even}, \\ p - a + 1 & i \text{ odd}; \end{cases}$$

and similarly for b_j. The placement for the various copies of X is indicated in the next figure.



Now let (x, y) be a cell in the gap between the copies of X in $P_{m,n}$, say, $x = p + 1$. Then $f(X)(x, y \pm 1) = 0$ as $(x, y \pm 1)$ also lies in the gap. By the definition of f we also have $f(X)(x+1, y) = f(X)(x-1, y)$. Therefore, $\sigma(f(X))(x, y) = 0$. Similarly

$$\sigma(f(X))(a_i + i \cdot (p+1), a_j + j \cdot (p+1)) = \sigma(X)(a, b)$$

for all $0 \leqslant i < m_0$, $0 \leqslant j < n_0$. Hence $\sigma(f(X)) = f(\sigma(X))$, showing that f is a simulation as claimed. Thus we have established the following lemma.

LEMMA 3. *For all $m, n \geqslant 1$ the kernel of rule σ on the $m \times n$ grid $P_{m,n}$ has dimension at least* $\gcd(m+1, n+1) - 1$. *In particular the $m \times n$ grid is irreversible under rule σ whenever $m+1$ and $n+1$ are not relatively prime.*

It remains to show that the lower bound of Lemma 3 is tight. To do this we must rule out the existence of configurations in the kernel of σ other than the ones obtained from the simulation just described. In the special case $n + 1 = k(m+1)$ for some $k \geqslant 1$ this follows immediately from (3.2): $d(P_{m,k(m+1)-1}) \geqslant \gcd(m+1, k(m+1)) - 1 = m$ by lemma 3 and $d(P_{m,k(m+1)-1}) \leqslant m = \min(m, n)$ by (3.2). The general situation can be handled as follows.

We proceed by induction on $\min(m, n)$. For $m = 1$ or $n = 1$ we are done by (2.3) as $P_{1,n}$ and $P_{n,1}$ are isomorphic to P_n. So assume that $\min(m, n) > 1$. As $P_{m,n}$ is isomorphic to $P_{n,m}$ we may assume that $n \geqslant m$. In the case $n = m$ we are done by the preceding remark, so assume $m < n$. Let $k := \lceil (n+1)/(m+1) \rceil$ and pick a

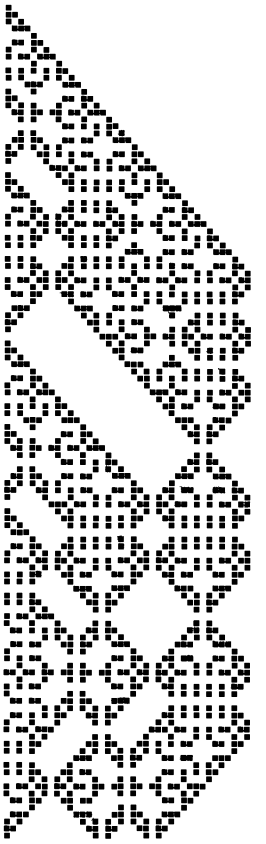


FIG. 3. The first half of configuration $Z_{1,39}^{+}$.

configuration X in the kernel of σ on $P_{m,n}$. Extend $Y_1 := X_1$ to a configuration in the kernel of σ on $P_{m,k(m+1)-1}$. The simulation f defined in (3.3) shows that

$$Y_j := \begin{cases} Y_{j \bmod (2m+2)} & \text{for } j \geqslant 2m+2 \text{ and} \\ Y_j = Y_{2m-j} & \text{for } m \leqslant j \leqslant 2m. \end{cases} \tag{3.4}$$

Here we assume for convenience that $Y_0 = \mathbf{0}$. But by proposition 2 X_1 can be extended in at most one way to a configuration in the kernel of σ on $P_{m,n'}$ for any n'. Hence $X_j = Y_j$ for all $j \leqslant n$. Furthermore, (3.4) implies that $X' := X_1 \cup X_2 \cup \cdots \cup X_{n'}$ lies in the kernel of σ on $P_{m,n'}$ where $n' := n$ **mod** $(2m+2)$ for $n \geqslant 2m+2$ and $n' := 2m - n$ otherwise. But $n' < m$, so by our induction hypothesis $d(P_{m,n'}) = \gcd(m+1, n'+1) - 1$. Now $\gcd(m+1, n'+1) = \gcd(m+1, n+1)$ and the theorem is proved. ■

Since $d(P_{m,n}) = \gcd(m+1, n+1) - 1$, it follows that every configuration in the kernel of $\sigma_{P_{m,n}}$ is of the form $f(X)$, where f: $P_{p,p} \to P_{m,n}$ is the simulation in (3.3), $p := \gcd(m+1, n+1) - 1$ and X is in the kernel of $\sigma_{P_{p,p}}$.

We conclude by pointing out the major difficulty with giving a similar analysis of the rule σ^{+}. The evolution of a one-point seed configuration on P_{∞} according to

m	$d^+(P_{m,m})$	m	$d^+(P_{m,m})$
4	4	53	2
5	2	54	4
9	8	59	22
11	6	61	40
14	4	62	24
16	8	64	28
17	2	65	42
19	16	67	32
23	14	69	8
24	4	71	14
29	10	74	4
30	20	77	2
32	20	79	64
33	16	83	6
34	4	84	12
35	6	89	10
39	32	92	20
41	2	94	4
44	4	95	62
47	30	98	20
49	8	99	16
50	8		

Irreversible σ^+-automata on square grids $P_{m,m}$, $m \leqslant 100$.

(3.1) for rule σ leads to a simple infinite checkerboard pattern. This is witnessed by the diamond shaped checkerboard configurations $Z_{k,m}$. The same evolution for rule σ^+, however, generates the pattern shown in Figure 2. The pattern is substantially more complicated, in fact, it has fractal dimension $\log_2(1 + \sqrt{17}) \approx 1.81$. If one defines configurations $Z^+_{k,m}$ analogously to the $Z_{k,m}$ configurations for rule σ the result looks typically like $Z^+_{1,39}$ as shown in our last figure. One can still prove that for every m there exists some $n \geqslant m$ such that $d^+(P_{m,n}) = m$, but the least such n appears to depend on number-theoretic properties of m in a rather complicated fashion. In particular we are not aware of any simple description of $d^+(P_{m,m})$. The following table contains the values of $d^+(P_{m,m})$ for $m \leqslant 100$. The missing entries are 0.

Acknowledgements. I would like to thank the referee for suggesting a number of improvements and pointing out several inaccuracies.

REFERENCES

1. E. F. Moore, Machine Models of Self-Reproduction, in: A. W. Burks, Essays on Cellular Automata, University of Illinois Press, 1970.
2. J. Myhill, The Converse of Moore's Garden-of-Eden Theorem, in: A. W. Burks, Essays on Cellular Automata, University of Illinois Press, 1970.
3. D. H. Pelletier, Merlin's magic square, *Amer. Math. Monthly*, 94 (1987) 143–150.
4. K. Sutner, Additive Automata on Graphs, submitted.
5. _____, Linear Cellular Automata and the Garden-of-Eden, to appear in *Math. Intelligencer*.
6. S. Wolfram, Geometry of binomial coefficients, *Amer. Math. Monthly*, 91 (1984) 566–571.
7. _____, Statistical mechanics and cellular automata, *Rev. Modern Physics*, 55 (1983) 601–644.
8. _____, Computer software in science and mathematics, *Scientific American*, 251 (1984) 188–203.

ACKNOWLEDGEMENTS

At this season, the Editors of the Monthly acknowledge with thanks the assistance of the referees, whose dedicated labor makes it all possible.

Arnold Adelberg, Jeanne Agnew, John Ahner, A. Aksoy, Michael Albertson, Leo Alex, Steven Althoen, William Anderson, Daniel Anderson, George Andrews, Thomas Angell, Tom Apostol, Ken Appel, Peter Ash, Robert Ash, Richard Askey, Ed Assmus, Kendall Atkinson, David Bailey, Robert Barnhill, Wayne Barrett, James Baumgartner, LeRoy Beasley, Gerald Beer, Donna Beers, Richard Beigel, Lowell Beineke, Howard Bell, Edward Bender, Martin Bendersky, John Benedetto, Russell Benson, Sterling Berberian, Simeon Berman, David Berman, Bruce Berndt, David Bindschadler, Garrett Birkhoff, D.M. Bloom, Ralph Boas, Kenneth Bogart, Richard Bollinger, Robert Borden, Peter Borwein, William Boyce, David Boyd, Bart Braden, Walter Brady, Mary Brewster William Briggs, David Browder, Herbert Brown, Morton Brown, Ezra Brown, Cynthia Brown, Richard Brualdi, T. Bruggeman, Creighton Buck, Edwin Buckman, Robert Burckel, S. Burr, David Burton

Greg Call, Frank Cannonito, Bille Carlson, George Carrier, Frank Carroll, C. Carroll, D. Castanon, Don Chakerian, Derek Chang, G. Chatterji, Arthur Chatters, Carmen Chicone, Lindsay Childs, Phyllis Chinn, Brong Choi, Joseph Cima, Barry Cipra, George Cobb, Todd Cochrane, Robert Cogburn, David Cohen, C. Coleman, Pierre Collet, William Cooke, Ken Cooke, Roger Cooke, E. Coven, J. Cremona, Bob Currier, Terry Czerwinski, Karl David, Kenneth Davidson, Philip Davis, Donald Dawson, Jane Day, Alice Dean, Carl DeBoor, Frank DeMeyer, William Derrick, Robert Devaney, Persi Diaconis, Beverly Diamond, Michael Drazin, Underwood Dudley, Richard Duffy, J. Dyer-Bennett

Hugh Edgar, Harold Edwards, Paul Eggermont, Leon Ehrenpreis, Jack Eidswick, R. Eldrikin, Jim Epperson John Ewing, Charles Eyndem, Jacek Fabrykowski, Neil Falkner David Farnsworth, Burton Fein, Michael Filaseta, Peter Fishburn, Stephen Fisher, Chris Fisher, Steve Fisk, Carl Fitzgerald, Harley Flanders, Wendell Fleming, David Ford, David Foulser, Michael Frame, Leroy Franklin, D. Freedman, Daniel Frohardt Wolfgang Fuchs, James Galambos, Joseph Gallian, Fred Galvin Betty Garrison, Susan Geller, I. Gentry, Murray Gerstenhaber, Peter Gibson, Richard Gillman, Andrew Gleason, Herman Gluck, David Gluck, Richard Goldberg, R. Goldman, Gianluca Gorni, S. Grabiner, James Graham-Eagle, David Grant, Andrew Granville J. Greever, Ralph Grimaldi, Carl Grimm, Emil Grosswald, Joanne Growney, Branko Grünbaum, William Gustafson

Jack Hale, D. Hall, Frank Hall, Katherine Halvorsen, Eugene Hamilton, David Harbater, William Hare, Theodore Harris, Donald Hartig Robin Hartshorne, Melvin Hauser, Michael Henle, Douglas Hensley David Herron, Gerald Hile, Theodore Hill, Robert Hill Peter Hilton, D. Hitt, Richard Hodel, Ross Honsberger Roger Horn, Tony Horowitz, Fredric Howard, George Hsiao, David Hull, James Hummel, John Huneke, Yasuhiko Ikebe, I. Isaacs Bernard Jacobson, Elliot Jacobson, Otomar Jajek, Robert James Thomas Jech, Charles Jepsen, Y. Jeung, Norman Johnson

Donald Kahn, D. Kalman, Wilfred Kaplan, Luise Kappe, Louis Kauffman, Nicholas Kazarinoff, Will Kazez, Paula Kemp, Judy Kennedy, E. Kerner, Jeanne Kerr, Irving Kessler, Harry Kesten, Bessie Kirkwood, James Kirkwood, David Klarner, Victor Klee, Arthur Knoebel, Neal Koblitz, Steve Kokoska Joseph Kung, Kenneth Kunin, Ray Kunze, Man Kwong, Walter Kyner Vincent La Riccia, Thomas Laffey, Jeffrey Lagarias, Gerard Lallement, Timothy Lance, Martin Landau, Loren Larson, Jimmie Lawson, Ronald Leach, Solomon Leader, John Leech Phyllis Lefton, David Legg, Derick Lehmer, Gerald Leibowitz Maita Levine, Lawrence Levy, Jonathan Lewin, James Lewis, Steve Ligh, Albert Livingston, Calvin Long, N. Lord

Thomas MacGregor, Robert MacLeod, Walter Malbury, John Mallet-Paret, Colin Mallows, Bennet Manvel, Daniel Marcus, Morris Marden, Tom Markham, Jerrold Marsden, M. Martelli Gary Martin, Peter Maserick, John McCleary, Richard McGehee Steve Medvedoff, Robert Melter, Jerry Metzger, Burnett Meyer, Kenneth Meyer, Ernest Michael, Kenneth Miller, David Minda Richard Mollin, Hugh Montgomery, Thomas Moore, Edward Moore John Morrison, James Munkries, Kent Nagle, Zuhair Nashid, Melvin Nathanson, D. Newman, Harold Niederreiter Albert Nijenhuis, Togo Nishiura, Ivan Niven, Katsumi Nomizu Robert Norton, Jeffrey Nunemacher, George O'Brien, Ingram Olkin, J. Ord, Howard Osborn, John Oxtoby, Frank Papp, Alan Parks, Beresford Parlett, Donald Peeples, Jaah Peetre, Doanld Pelletier, David Penney, Stephen Penrice, Robert Phelps, Richard Pierce, A. Pixley, Carl Pomerance, Robert Powers

J. Quine, Dahiya Raj, Maxwell Reade, David Reiner, Jeff Remmel, Michael Rennie, Bruce Reznick, Billy Rhoadis, L. Ribes, David Richman, Bill Rising, Theodore Rivlin, David Roedu, Bob Roos David Roselle, Michael Rosen, Joseph Rotman, Ranjan Roy W. Ruckle, Walter Rudin, Patrick Ryan, E. Saff, Bruce Sagan, Hans Sagan, Yoram Sagher, Donald Sarason, Andre Scedrov, Glenn Schober, John Schue, H. Schultz, Berthold Schweizer Richard Segall, John Selfridge, Marjorie Senechal, Robert Serfling, Louis Shapiro, David Shmoys, Yasutaka Sibuya, Herb Silverman, Eveyln Silvia, Benjamin Sims, David Singmaster, Richard Sinkhorn, Abe Sklar, Peter Slater, C. Small, Don Small, David Smith, Craig Smorynski, Laurie Snell, Lawrence Somer, Jerome Spanier, David Spring, Richard Stanley, Norton Starr, Sherman Stein, C. Stewart, B. Stewart, Gilbert Strang, Karl Stromberg, David Styer, Ted Suffridge, Richard Swan, Charles Swartz, John Sylvester

Elliot Tanis, Christopher Thompson, Brian Thomson, G. Trapp, Janet Tremain, Anne Trembinska, William Trench, Thomas Tucker, Alan Tucker, N. Valiaho, M. Van Rossum, Scott Vanstone, Ilan Vardi, Richard Varga, Dan Vellemin, Rudolf Vyborny, Michelle Wachs Samuel Wagstaff, Martin Walter, Derming Wang, William Wardlaw, Mary Wardrop-Embry, William Waterhouse, William Webb, Cliff Weil, Alan Weinstein, John Wetzel, Dennis White, Donald Wilken, James Yorke, Anne Young, E. Zachmanoglou, Dan Zelinsky, David Zwick

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief descripition of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Alberta, Canada T2N 1N4.

More on Mortality

MELVEN KROM
Mathematics Department, University of California, Davis, CA 95616

MYREN KROM
Mathematics Department, California State University, Sacramento, CA 95819

A set of matrices is called ***mortal*** if the zero matrix can be expressed as a product of members of the set. For example,

$$\left\{\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix},\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\right\}$$

is not mortal but

$$\left\{\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix},\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\right\}$$

is mortal because

$$\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}.$$

Question 1. Does there exist an algorithm which decides for any finite set of 2×2 matrices with integer entries whether the set is mortal?

It is easy to find examples of mortal sets and examples of nonmortal sets of matrices. A natural question to ask is whether there is an algorithm to determine mortality. You might be able to answer the question in terms of basic properties of matrices and of integers by discovering, from examples, a systematic way of checking for mortality. Indeed, if we limit the question to lower triangular matrices, there is a simple algorithm to determine mortality. A finite set of lower triangular 2×2 matrices is mortal if and only if, for each diagonal entry position there is a matrix in the set with a zero in that position.

It is important to realize that there may not exist an algorithm as indicated in question 1. That is, there may not be a recursive procedure [**1**, p. 534], [**2**] for deciding mortality and we would say that the mortality problem is unsolvable [**4**, p. 2], [**2**]. It is possible to use elementary means and a known result that some other problem is recursively unsolvable to verify that a mortality problem is unsolvable. This was done in [**5**] to show that the mortality problem for 3×3 matrices is unsolvable. Thus, if we change question 1 to refer to 3×3 matrices instead of 2×2

matrices the answer to the question is no. The proof in [5] of the unsolvability of the 3×3 case consists of showing how to form, for each instance of a known unsolvable decision problem, a finite set of 3×3 matrices that is mortal if and only if the answer to the instance of the unsolvable decision problem is yes. For examples of known unsolvable problems that might be used in this way for the 2×2 case, see the many references given in [2].

Question 1 was first mentioned as an open problem in [6] where it is discussed in terms of transformations of the complex plane. In [3] question 1 is shown to be equivalent to question 2 below. Question 2 is a reformulation of the mortality problem as an "equality of entry" problem. We let c_{ij} denote the entry in the ith row and jth column of the matrix C.

Question 2. Does there exist an algorithm which decides for any finite set P of nonsingular 2×2 matrices with integer entries whether there is a product that is formed with members of P that is equal to a matrix C such that $c_{21} = c_{22}$?

If we approach the mortality problem, question 1, by restricting it to lower triangular matrices then the problem becomes trivially solvable. But if we approach the equivalent equality of entry problem, question 2, by restricting it to lower triangular matrices then it becomes a nontrivial open problem, question 3.

Question 3. Does there exist an algorithm which decides for any finite set P of nonsingular lower triangular 2×2 matrices with integer entries whether there is a product that is formed with members of P that is equal to a matrix C such that $c_{21} = c_{22}$?

REFERENCES

1. J. Barwise, editor, Handbook of Mathematical Logic, North-Holland, Amsterdam.
2. J. P. Jones, Recursive undecidability—an exposition, this MONTHLY, 81 (1974) 724–738.
3. Melven Krom and Myren Krom, Recursive solvability of problems with matrices, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, to appear.
4. H. R. Lewis, Unsolvable Classes of Quantificational Formulas, Addison-Wesley, Reading, 1979.
5. M. S. Paterson, Unsolvability in 3×3 matrices, *Studies in Appl. Math.*, 49 (1970) 105–107.
6. P. Schultz, Mortality of 2×2 matrices, this MONTHLY, 84 (1977) 463–464; correction, 85 (1978) 263.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

Elementary Evaluations of $\int_0^\infty e^{-x^2}\,dx$, $\int_0^\infty \cos x^2\,dx$, and $\int_0^\infty \sin x^2\,dx$

ROBERT WEINSTOCK
Department of Physics, Oberlin College, Oberlin, OH 44074

Less than a decade ago [1] there appeared a new method for evaluating

$$J = \int_0^\infty e^{-x^2}\,dx, \tag{1}$$

long known to have the value $\sqrt{\pi}/2$. The cleverness of the method resides in the introduction of

$$e(x) = \int_{-\infty}^{\infty} \frac{e^{-x(1+t^2)}}{1+t^2}\,dt, \tag{2}$$

whose differentiation leads readily to $4J^2 = \pi$. Since $e(x)$ is defined by an improper integral, however, one cannot use the elementary Leibniz rule to compute its derivative: One must either apply a deeper theorem—e.g., [2]—or perform an *ad hoc* computation to form $e'(x)$, as is done in [1].

The present note offers a modification of the method in [1] that borrows the main item from [1], but which avoids the immanent complication of having to differentiate an improper integral.

One finds in [1] the evaluation of the Fresnel integrals by essentially the same method, *mutatis mutandis*, that is used there to evaluate (1); again an awkwardness intrudes through the need to differentiate improper integrals. The present note modifies the method so as to evaluate the Fresnel integrals while avoiding such complications.

First, the evaluation of (1): In place of (2) we define

$$f(x) = \int_0^1 \frac{e^{-x(1+t^2)}}{1+t^2}\,dt \tag{3}$$

and note—along with the fact that $f(x)$ is continuous and differentiable for all real x—that

$$f(0) = \arctan 1 = (\pi/4), \qquad f(\infty) = 0; \tag{4}$$

the latter because, when $x > 0$,

$$0 < f(x) = e^{-x}\int_0^1 \frac{e^{-xt^2}}{1+t^2}\,dt < e^{-x}(\pi/4).$$

Application of the Leibniz rule to (3) gives, for $x > 0$,

$$f'(x) = -\int_0^1 e^{-x(1+t^2)}\,dt = -e^{-x}\int_0^1 e^{-xt^2}\,dt = -\frac{e^{-x}}{\sqrt{x}}\int_0^{\sqrt{x}} e^{-u^2}\,du = -\frac{e^{-x}}{\sqrt{x}}g(\sqrt{x}), \tag{5}$$

where

$$g(z) = \int_0^z e^{-u^2}\,du. \tag{6}$$

Integrating (5) from 0 to ∞, then making the substitution $z = \sqrt{x}$, we use (6) to obtain

$$\begin{aligned} f(\infty) - f(0) &= -\int_0^\infty \frac{e^{-x}}{\sqrt{x}} g(\sqrt{x})\,dx = -2\int_0^\infty e^{-z^2} g(z)\,dz \\ &= -2\int_0^\infty g'(z)g(z)\,dz = [g(0)]^2 - [g(\infty)]^2. \end{aligned} \tag{7}$$

Because of (1), (4), and (6), we read (7) as $-(\pi/4) = -J^2$—i.e., $J = \sqrt{\pi}/2$.

For evaluation of the Fresnel integrals

$$F = \int_0^\infty \cos y^2\,dy, \qquad G = \int_0^\infty \sin y^2\,dy, \tag{8}$$

we introduce, in analogy to (3),

$$\alpha(x) = \int_0^1 \frac{\cos xt^2}{1+t^2}\,dt, \qquad \beta(x) = \int_0^1 \frac{\sin xt^2}{1+t^2}\,dt. \tag{9}$$

We note—along with the fact that $\alpha(x)$ and $\beta(x)$ are continuous and differentiable for all real x—that

$$\alpha(0) = \frac{\pi}{4}, \qquad \beta(0) = 0. \tag{10}$$

Furthermore, we show below that

$$\alpha(\infty) = 0, \qquad \beta(\infty) = 0. \tag{11}$$

Applying the Leibniz rule to both integrals in (9), we use algebraic rearrangement to obtain a pair of linear differential equations

$$\alpha'(x) - \beta(x) = -q(x), \qquad \beta'(x) + \alpha(x) = p(x), \tag{12}$$

where, for $x > 0$,

$$\left.\begin{aligned} p(x) &= \int_0^1 \cos xt^2\,dt = \frac{1}{\sqrt{x}}\int_0^{\sqrt{x}} \cos y^2\,dy = \frac{1}{\sqrt{x}}u(\sqrt{x}) \\ q(x) &= \int_0^1 \sin xt^2\,dt = \frac{1}{\sqrt{x}}\int_0^{\sqrt{x}} \sin y^2\,dy = \frac{1}{\sqrt{x}}v(\sqrt{x}) \end{aligned}\right\}, \tag{13}$$

and,

$$u(z) = \int_0^z \cos y^2\,dy, \qquad v(z) = \int_0^z \sin y^2\,dy. \tag{14}$$

We treat (12) as follows: (i) multiply the first by $-\cos x$, the second by $\sin x$, then add the products; (ii) multiply the first by $\sin x$, the second by $\cos x$, then add.

The two results can be written, respectively,

$$\left.\begin{aligned} &\text{(i)} \quad \frac{d}{dx}\{\beta(x)\sin x - \alpha(x)\cos x\} = p(x)\sin x + q(x)\cos x \\ &\text{(ii)} \quad \frac{d}{dx}\{\beta(x)\cos x + \alpha(x)\sin x\} = p(x)\cos x - q(x)\sin x \end{aligned}\right\}. \tag{15}$$

Integrating (15, i) and (15, ii) from 0 to ∞, we obtain, on application of (10) and (11),

$$\int_0^\infty [p(x)\sin x + q(x)\cos x]\,dx = \frac{\pi}{4}, \qquad \int_0^\infty [p(x)\cos x - q(x)\sin x]\,dx = 0. \tag{16}$$

Now, using (13), the substitution $z = \sqrt{x}$, and (14), we rewrite (16) as

$$\begin{aligned} \frac{\pi}{4} &= 2\int_0^\infty [u(z)\sin z^2 + v(z)\cos z^2]\,dz = 2\int_0^\infty [u(z)v'(z) + v(z)u'(z)]\,dz \\ &= 2\int_0^\infty \frac{d}{dz}[u(z)v(z)]\,dz = 2FG, \end{aligned} \tag{17}$$

$$\begin{aligned} 0 &= 2\int_0^\infty [u(z)\cos z^2 - v(z)\sin z^2]\,dz = 2\int_0^\infty [u(z)u'(z) - v(z)v'(z)]\,dz \\ &= \int_0^\infty \frac{d}{dz}\{[u(z)]^2 - [v(z)]^2\}\,dz = F^2 - G^2 \end{aligned} \tag{18}$$

—with further help, in both cases, from (8) and (14).

To show that G is positive (and with it F, by (17)), one substitutes $y = \sqrt{x}$ in (8) and writes

$$G = \frac{1}{2}\int_0^\infty \frac{\sin x}{\sqrt{x}}\,dx = \frac{1}{2}\sum_{k=0}^{\infty}\int_{\pi k}^{\pi(k+1)} \frac{\sin x}{\sqrt{x}}\,dx$$

—obviously an alternating series of diminishing terms whose k^{th} term approaches 0 as $k \to \infty$ and whose $k = 0$ term is positive. It follows that $G > 0$. From (17) and (18), therefore, $F = G = (1/2)\sqrt{\pi}/2$.

To prove the second in (11), we substitute $y = xt^2$ in (9) $(x > \pi)$

$$\beta(x) = \frac{1}{2\sqrt{x}}\int_0^x \frac{\sin y}{\sqrt{y}\,[1 + (y/x)]}\,dy. \tag{19}$$

We partition $[0, x]$ into the subintervals $[0, \pi], [\pi, 2\pi], \ldots, [n\pi, x]$ for the appropriate positive integer n, and write (19) as a sum of integrals over the subintervals; this obviously results in a sum of diminishing alternating-sign terms, of which the first is positive. Thus, for $x > \pi$,

$$0 < \beta(x) < \frac{1}{2\sqrt{x}}\int_0^\pi \frac{\sin y}{\sqrt{y}\,[1 + (y/x)]}\,dy < \frac{1}{2\sqrt{x}}\int_0^\pi \frac{\sin y}{\sqrt{y}}\,dy,$$

whence $\beta(\infty) = 0$.

It is almost as obvious that a like partition of the interval $[0, x]$ for evaluation of

$$\alpha(x) = \frac{1}{2\sqrt{x}} \int_0^x \frac{\cos y}{\sqrt{y}\,[1 + (y/x)]}\, dy$$

leads to the following bracketing of $\alpha(x)$ when $x > \pi$:

$$0 < \alpha(x) < \frac{1}{2\sqrt{x}} \int_0^{\pi} \frac{\cos y}{\sqrt{y}\,[1 + (y/x)]}\, dy$$

$$< \frac{1}{2\sqrt{x}} \int_0^{\pi/2} \frac{\cos y}{\sqrt{y}\,[1 + (y/x)]}\, dy < \frac{1}{2\sqrt{x}} \int_0^{\pi/2} \frac{\cos y}{\sqrt{y}}\, dy$$

—whence, also, $\alpha(\infty) = 0$. This completes the proof of (11).

A remark on personal motivation. The use of 1 as upper integration limit in (3) was not my original choice: Given (2) and wishing to avoid differentiating an infinite integral in a classroom presentation of the method of [1], I tried using arbitrary $r > 0$ as upper limit in (3), with the intention of $r \to \infty$ eventually. The device worked quite simply; but I later noticed that setting $r = 1$ makes the calculation simpler yet.

Thanks go to the (second) referee for suggesting that I try to evaluate the Fresnel integrals (8) by means of a finite-integral modification of the device used in [1], and for calling my attention to yet another recent evaluation of them [3]. (Only J was evaluated in the note I first submitted.)

REFERENCES

1. J. van Yzeren, Moivre's and Fresnel's integrals by simple integration, this MONTHLY, 86 (1979) 691–693.
2. T. M. Apostol, Mathematical Analysis, Addison-Wesley, 1957, p. 443.
3. H. Flanders, On the Fresnel integrals, this MONTHLY, 89 (1982) 264–266.

The L^4 Norm of a Polynomial with Coefficients ± 1

DONALD J. NEWMAN AND J. S. BYRNES,
Prometheus, Inc., 103 Mansfield Street, Sharon, MA 02067

A classic unresolved question regarding polynomials

$$P(z) = \sum_{k=0}^{n-1} \varepsilon_k(n) z^k$$

with coefficients $\varepsilon_{k(n)} = \varepsilon_k = \pm 1$ is whether the maximum modulus of such a polynomial on the unit circle can be $n^{1/2} + o(n^{1/2})$. As shown by Kahane [3], if complex coefficients of modulus 1 are allowed, then not only is it possible for this property to be satisfied, but the minimum modulus can be $n^{1/2} + o(n^{1/2})$ as well. Specifically, Kahane proved that for any n there is a polynomial of degree n with coefficients of modulus one whose modulus everywhere on the unit circle is $n^{1/2} + o(n^{3/10} \log n)$.

Erdös [2] had conjectured the existence of a $c > 0$ such that, for any polynomial P of the types described, $\|P\|_\infty > (1 + c)n^{1/2}$. Clearly the Kahane result disproved this conjecture for the modulus 1 case, but the situation for coefficients ± 1 remains open. Employing an elegant construction Shapiro [6, 5] demonstrated the achievability of the order of magnitude $n^{1/2}$, but the maximum modulus of the Shapiro polynomials is about $(2n)^{1/2}$.

Since the L^2 norm of any such polynomial is $n^{1/2}$, by the Parseval Theorem, and the L^P norm for any $p < \infty$ is a lower bound for the maximum modulus, it is natural to look at the L^P norm for some $p > 2$. Motivated by these considerations, as well as the inherent tractability of the L^4 norm, we examine the L^4 norm of such polynomials. As one might expect, this leads to several interesting combinatorial questions. We provide answers to some of these, and conclude with a refined version of the Erdös conjecture.

Throughout the note n will be a positive integer, $P(z)$ will denote the previously indicated polynomial of degree $n - 1$ with coefficients ± 1, and z will lie on the unit circle, $z = e^{2\pi i\theta}$, $0 \leqslant \theta < 1$. All integrals will be over $\theta \in [0, 1]$. We begin with a lemma.

LEMMA.

$$\|P\|_{L^4}^4 = \sum_{\substack{j+k=l+m \\ 0\leqslant j,\, k,\, l,\, m<n}} \varepsilon_j\varepsilon_k\varepsilon_l\varepsilon_m.$$

Proof.

$$\begin{aligned}\|P\|_{L^4}^4 &= \int_0^1 \left|P\left(e^{2\pi i\theta}\right)\right|^4 d\theta \\ &= \int \left(\sum_{j=0}^{n-1} \varepsilon_j z^j\right)\left(\sum_{k=0}^{n-1} \varepsilon_k z^k\right)\left(\sum_{l=0}^{n-1} \varepsilon_l z^{-l}\right)\left(\sum_{m=0}^{n-1} \varepsilon_m z^{-m}\right) \\ &= \int \text{constant terms}.\end{aligned}$$

Since a constant term occurs in this product if and only if $j + k = l + m$, the result follows immediately.

Of interest is the expected value $E(\|P\|_{L^4}^4)$, if the coefficients ε_j are chosen at random.

THEOREM 1. $E(\|P\|_{L^4}^4) = 2n^2 - n$.

Proof. Clearly if exactly 3 of the indices j, k, l, m are identical, or if at least 3 of them are different, then $E(\varepsilon_j \cdot \varepsilon_k \cdot \varepsilon_l \cdot \varepsilon_m) = 0$. Therefore, it follows from the lemma that

$$E\left(\|P\|_{L^4}^4\right) = \sum_{\substack{j=l \text{ and } k=m \\ \text{or } j=m \text{ and } k=l}} \varepsilon_j\varepsilon_k\varepsilon_l\varepsilon_m. \tag{1}$$

For each of the $n(n-1)/2$ pairs of integers p, q, $0 \leqslant p < q < n$, there are 4 terms appearing in (1), $\varepsilon_p\varepsilon_q\varepsilon_p\varepsilon_q$ being typical. The only other terms in (1) are ε_p^4,

$0 \leqslant p < n$. Since all of these terms equal 1,

$$E\left(\|P\|_{L^4}^4\right) = n + 4(n)(n-1)/2 = 2n^2 - n,$$

completing the proof of Theorem 1.

We now observe the improvement that is achieved when this random choice is replaced by the Shapiro coefficients. Shapiro's polynomials are defined, together with his auxiliary polynomials Q, by the recurrence formulas

$$\begin{aligned} P_0(z) &= Q_0(z) = 1, \\ P_{m+1}(z) &= P_m(z) + z^{2^m} Q_m(z), \\ Q_{m+1}(z) &= P_m(z) - z^{2^m} Q_m(z), \qquad m \geqslant 0. \end{aligned} \tag{2}$$

For the interested reader we point out articles by Brillhart et al. [1] and Mendès-France and Tenenbaum [4]. The former shows that the Shapiro coefficients can be defined directly from the binary representation of the order of the polynomial, while the latter relates them to paper-folding sequences.

THEOREM 2. *If $n = 2^k$ and $P(z)$ is the Shapiro polynomial of degree $n-1$, then*

$$\|P\|_{L^4}^4 = \left(4n^2 - (-1)^k n\right)/3.$$

Proof. It follows directly from (2) that

$$|P_{m+1}(z)|^2 + |Q_{m+1}(z)|^2 = 2\left(|P_m(z)|^2 + |Q_m(z)|^2\right)$$

so that, as observed by Shapiro,

$$|P_m(z)|^2 + |Q_m(z)|^2 \equiv 2^{m+1}. \tag{3}$$

Now (2) and (3) yield

$$|P_{m+1}(z)|^2 = 2^{m+1} + 2\,\mathrm{Re}\left(z^{2^m} Q_m(z)\overline{P}_m(z)\right). \tag{4}$$

Next we observe that $z^{2^m} Q_m \overline{P}_m$ is composed solely of frequencies which are positive powers of z, so that it can be thought of as $Q_m \tilde{P}_m$, where $\tilde{P}_m$ is the "reversed" polynomial of P. Thus

$$\int |P_{m+1}|^4 = 2^{2m+2} + 4\int \left(\mathrm{Re}\, Q_m \tilde{P}_m\right)^2.$$

Since $f \equiv Q\tilde{P}$ is analytic and 0 at the origin,

$$0 = \int f^2 = \mathrm{Re}\left(\int f^2\right) = \int (\mathrm{Re}\, f)^2 - \int (\mathrm{Im}\, f)^2,$$

so that

$$\int (\mathrm{Re}\, f)^2 = \tfrac{1}{2}\int |f|^2 = \tfrac{1}{2}\int |Q_m|^2 |\tilde{P}_m|^2 = \tfrac{1}{2}\int |P_m|^2 |Q_m|^2.$$

Altogether then, we have

$$\begin{aligned} \int |P_{m+1}|^4 &= 2^{2m+2} + 2\int |P_m|^2 \left(2^{m+1} - |P_m|^2\right) \\ &= 2^{2m+3} - 2\int |P_m|^4. \end{aligned} \tag{5}$$

The remainder of the proof is now simply induction on k. The result is obviously true for $k = 0$, since $P_0(z) = 1$. Furthermore, from (5) and the inductive hypothesis

$$\int |P_k|^4 = \frac{2^{2k+2} - (-2)^k}{3},$$

it follows that

$$\int |P_{k+1}|^4 = 2^{2k+3} - 2\frac{2^{2k+2} - (-2)^k}{3} = \frac{2^{2(k+1)+2} - (-2)^{k+1}}{3},$$

as required. This completes the proof of Theorem 2.

Note that Theorem 2 implies that the L^4 norm of the $(n-1)$st-degree Shapiro polynomial is asymptotic to $\sqrt{n}$ times the fourth root of $4/3 \approx 1.07457\sqrt{n}$. Based upon extensive numerical evidence employing the Bose-Einstein statistics methodology of statistical mechanics, we conjecture that the Shapiro polynomials do not give the minimum L^4 norm among all polynomials of the same degree with coefficients ± 1, but that this minimum L^4 norm is asymptotically $\sqrt{n}$ times the fourth root of $6/5 \approx 1.04664\sqrt{n}$. Observe that the truth of this conjecture would imply that of the Erdös conjecture mentioned earlier, with $c = (6/5)^{1/4} - 1 \approx .04664$.

Research sponsored by the Air Force Office of Scientific Research (AFSC), under Contract F49620-86-C-0088.

REFERENCES

1. John Brillhart, J. S. Lomont, and Patrick Morton, Cyclotomic properties of the Rudin-Shapiro polynomials, *J. Reine Angew. Math.*, 288 (1976) 37–65.
2. Paul Erdös, Some unsolved problems, *Mich. Math. J.*, 4 (1957) 291–300.
3. Jean-Pierre Kahane, Sur les Polynomes à Coefficients Unimodulaires, *Bull. London Math Soc.*, 12 (1980) 321–342.
4. Michel Mendès-France and Gerald Tenenbaum, Dimension des Courbes Planes, Papiers Plies et Suite de Rudin-Shapiro, *Bull. Soc. Math. France*, 109 (1981) no. 2, 207–215.
5. W. Rudin, Some theorems on Fourier coefficients, *Proc. Amer. Math. Soc.*, 10 (1959) 855–859.
6. H. S. Shapiro, Extremal Problems for Polynomials and Power Series, thesis, Mass. Inst. of Tech. (1957).

Isoptic Characterization of a Circle (Proof of a Conjecture of M. S. Klamkin)

JOHANNES C. C. NITSCHE
Department of Mathematics, University of Minnesota, Minneapolis, MN 55455

Suppose that a smooth convex curve C has the property that as we walk around a certain circle K that contains C in its interior, the curve C always subtends the same angle. Can we conclude that C is a circle? We cannot [1], but Murray Klamkin [2] has conjectured that if there are two concentric circles K_1 and K_2 with respect to which C has this "isoptic" property, then C must indeed be a circle. In this note we prove his conjecture.

The locus of intersection of tangents to a smooth convex curve C meeting at a constant angle, say $\pi - \beta$, is an *isoptic* of C; if $\beta = \pi/2$, the isoptic is called an *orthoptic*.

Isoptics of a circle are circles. As J. W. Green [1] has shown, the converse of this statement is only true if β is an irrational multiple of π or if $\beta = (m/n)\pi$, where m and n are relatively prime positive integers and m is even. For example, the circle $x^2 + y^2 = a^2 + b^2$ is an isoptic for the ellipse $(x/a)^2 + (y/b)^2 = 1$ corresponding to the angle $\beta = \pi/2$.[1]

Let C have two concentric isoptic circles of radii R_1 and $R_2 > R_1$, with center at the origin, and corresponding angles $\beta_1 = (m_1/n_1)\pi$, $\beta_2 = (m_2/n_2)\pi$, where $0 < \beta_1 < \beta_2 < \pi$. For $j = 1, 2$, the integers m_j and n_j are relatively prime, and m_j is odd. It is easily verified that the origin must be an interior point of C. Denote by $h(\theta)$ the Minkowski supporting function of C, chosen with respect to the origin. This function measures the distance between the origin and the unique oriented tangent to C which forms the angle $\theta + \pi/2$ with the positive x-axis. Its equation is $x\cos\theta + y\sin\theta = h(\theta)$. The curve C appears as the envelope of its tangents and can be represented in the form $x = h(\theta)\cos\theta - h'(\theta)\sin\theta$, $y = h(\theta)\sin\theta + h'(\theta)\cos\theta$. The curvature of C turns out to be $[h(\theta) + h''(\theta)]^{-1}$.

We have $0 < h(\theta) < R_1$, and the following relations are read off from Figure 1:

$$\operatorname{Arc}\cos\left[h(\theta)/R_j\right] + \operatorname{Arc}\cos\left[h(\theta + \beta_j)/R_j\right] = \beta_j, \qquad j = 1, 2. \tag{1}$$

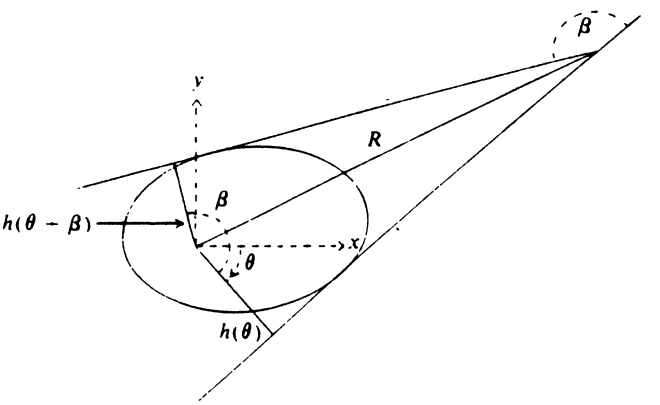


FIG. 1.

Here Arc cos stands for the principal branch of the inverse function $\cos^{-1}$. Advancing the argument in (1) by β_j and subtracting the resulting relation leads to $h(\theta + 2\beta_j) = h(\theta)$. Thus $h(\theta)$ has the periods $2\beta_1$ and $2\beta_2$, in addition to the

[1] Thus, a denizen of *Flatland* in Edwin Abbott's charming story, walking by chance around this circle, would be unable to distinguish the shape of the ellipse from that of the circle $x^2 + y^2 = (a^2 + b^2)/2$, because both would appear to him consistently under the visual angle $\pi/2$. If this denizen were enlightened, i.e., cognizant of our theorem, he could settle the question by repeating his angular measurements from the points of a second circle.

period 2π. There are integers a_j, b_j for which $a_j m_j + b_j n_j = 1$, so that

$$a_j 2\beta_j + b_j 2\pi = a(m_j/n_j)2\pi + b_j 2\pi = 2\pi(a_j m_j + b_j n_j)/n_j = 2\pi/n_j$$

are periods of $h(\theta)$. As the m_j are odd, then $h(\theta + \beta_j) = h(\theta + \pi/n_j)$, and relation (1) becomes

$$\operatorname{Arc}\cos\left[h(\theta)/R_j\right] + \operatorname{Arc}\cos\left[h\left(\theta + \pi/n_j\right)/R_j\right] = \beta_j, \qquad j = 1,2 \tag{2}$$

or, equivalently,

$$h^2(\theta) + h^2\left(\theta + \pi/n_j\right) - 2\cos\beta_j h(\theta)h\left(\theta + \pi/n_j\right) = R_j^2 \sin^2\beta_j, \qquad j = 1,2. \tag{3}$$

Let $d = (n_1, n_2)$ be the g.c.d. of the denominators and set $n_j = \nu_j d$. There are integers a and b for which $b\nu_1 + a\nu_2 = 1$, so that $(2\pi/n_1)a + (2\pi/n_2)b = 2\pi(a\nu_2 + b\nu_1)/d\nu_1\nu_2 = 2\pi/N$, with $N = \nu_1\nu_2 d = n_1\nu_2 = n_2\nu_1$. Therefore the number $p = 2\pi/N$ is a period of the supporting function $h(\theta)$.

Observe that $\pi/n_1 = p\nu_2/2$ and $\pi/n_2 = p\nu_1/2$. It follows that $h(\theta + \pi/n_1) = h(\theta)$ if ν_2 is even and $h(\theta + \pi/n_2) = h(\theta)$ if ν_1 is even. In either case, (2) implies that $h(\theta) = \text{const.}$, i.e., that C must be a circle. If $\nu_1 = 2k_1 + 1$ and $\nu_2 = 2k_2 + 1$, then $\pi/n_1 = k_2 p + p/2$, $\pi/n_2 = k_1 p + p/2$, and subtraction of the two relations (3) gives

$$2[\cos\beta_1 - \cos\beta_2]h(\theta)h(\theta + \pi/N) = R_2^2 \sin^2\beta_2 - R_1^2 \sin^2\beta_1$$

or, since $0 < \beta_1 < \beta_2 < \pi$, $h(\theta)h(\theta + \pi/N) = \text{const.}$ In conjunction with (3), this fact implies $h(\theta) = \text{const.}$, so that, again, C must be a circle.

Remark. It might be of interest to consider similar questions in higher dimensions.

REFERENCES

1. J. W. Green, Sets subtending a constant angle on a circle, *Duke Math. J.*, 17 (1950) 263–267.
2 M. S. Klamkin, Conjectured isoptic characterization of a circle, *Amer. Math. Monthly*, 95 (1988) 845.

Two Theorems on Singular Values and Eigenvalues

ROY MATHIAS
Department of Mathematical Sciences, Johns Hopkins University, Baltimore, MD 21218

We present elementary proofs, based on the eigenvalue interlacing property of Hermitian matrices, of two theorems relating the singular values and eigenvalues of a square complex matrix.

Let $M_{m,n}$ denote the set of n-by-m complex matrices and set $M_n \equiv M_{n,n}$. For a given $A = [a_{ij}] \in M_n$, denote its eigenvalues by $\{\lambda_i(A)\}$, always ordered in decreasing modulus

$$|\lambda_1(A)| \geq |\lambda_2(A)| \geq \cdots \geq |\lambda_n(A)|.$$

The function $\rho(A) \equiv |\lambda_1(A)|$ is the *spectral radius* of A.

We define the *singular values* of $A \in M_{m,n}$ by $\sigma_i(A) \equiv [\lambda_i(A^*A)]^{1/2}$, $i = 1, \ldots, \min\{m, n\}$. Note that A^*A is positive semi-definite and because of our ordering convention for eigenvalues,

$$\sigma_1(A) \geqslant \sigma_2(A) \geqslant \cdots \geqslant \sigma_{\min\{m,n\}}(A) \geqslant 0.$$

Since the eigenvalues of a matrix are a similarity invariant, it follows that the singular values of a matrix are invariant under unitary equivalence, i.e., for any unitary U and V,

$$\begin{aligned}\sigma_i(UAV) &= \left[\lambda_i((UAV)^*(UAV))\right]^{1/2} = \left[\lambda_i(V^*A^*AV)\right]^{1/2} \\ &= \left[\lambda_i(A^*A)\right]^{1/2} = \sigma_i(A).\end{aligned}$$

Notice that if $A \in M_n$ is singular, then so is A^*A, and hence $0 = [\lambda_n(A^*A)]^{1/2} = \sigma_n(A)$. The function $|||A|||_2 \equiv \sigma_1(A)$ is called the *spectral norm* of A; it is the operator norm of A induced by the usual Euclidean norm $\|x\|_2 = (x^*x)^{1/2}$ on $\mathbb{C}^n$, i.e., $|||A|||_2 = \max\{\|Ax\|_2 : x \in \mathbb{C}^n, \|x\|_2 = 1\}$.

There are two well-known relations between the spectral norm and the spectral radius:

$$|||A|||_2 \geqslant \rho(A) \tag{1}$$

$$\lim_{m \to \infty} |||A^m|||_2^{1/m} = \rho(A). \tag{2}$$

Both are easily proved, and both are true more generally for any matrix norm, not just the spectral norm (see Chapter 5 of [3]).

We are interested in two generalizations of these relations that involve eigenvalues and singular values other than the largest.

THEOREM 1 (Weyl [5]). *Let $A \in M_n$. Then*

$$\sigma_1(A)\sigma_2(A) \cdots \sigma_k(A) \geqslant |\lambda_1(A)\lambda_2(A) \cdots \lambda_k(A)| \quad \text{for } k = 1, \ldots, n. \tag{3}$$

THEOREM 2 (Yamamoto [6]). *Let $A \in M_n$. Then*

$$\lim_{m \to \infty} \left[\sigma_k(A^m)\right]^{1/m} = |\lambda_k(A)| \quad \text{for } k = 1, 2, \ldots, n. \tag{4}$$

Notice that the case $k = 1$ in these theorems is just (1) and (2), and that the inequality in Theorem 1 is actually an equality for $k = n$ since both sides are equal to $|\det A|$ in this case. Consider the case $k = n$ in (4): If A is singular, then so is A^m and $\sigma_n(A^m) = \lambda_n(A^m) = 0$ for every $m = 1, 2, \ldots$. If A is nonsingular, the assertion follows from applying (2) to A^{-1}, since $\sigma_1(A^{-1}) = 1/\sigma_n(A)$ and $\lambda_1(A^{-1}) = 1/\lambda_n(A)$.

These two theorems can be deduced easily from (1) and (2), respectively, by using the notion of a *compound matrix*, and this is how they are proved in [5] and [6]. Other proofs of Theorem 1 are known that do not involve the compound, e.g., [2] and Lemma 3.3 in Chapter 2 of [1], but there seems to be no other proof of Theorem 2 available. We shall show that both theorems are easily deduced from Lemma 3, which follows directly from the interlacing property for eigenvalues of a Hermitian matrix.

LEMMA 3 (Interlacing property of singular values). *Let $A \in M_n$ be given, and let $A_p \in M_{n,n-p}$ (respectively, $A_p \in M_{n-p,n}$) denote a submatrix of A obtained by*

deleting any p columns (respectively, any p rows) from A. Then

$$\sigma_i(A) \geqslant \sigma_i(A_p) \geqslant \sigma_{i+p}(A) \quad \text{for } i = 1, 2, \ldots, n-p. \tag{5}$$

Although deleting rows or columns of a matrix may decrease its singular values, the next lemma shows that this does not happen when zero rows or columns are deleted.

LEMMA 4. *Let $A \in M_n$ be given, and consider $B = [0|A] \in M_{n,n+p}$ and $C = \begin{bmatrix} A \\ 0 \end{bmatrix} \in M_{n+p,n}$ obtained by adjoining p columns (respectively, p rows) of zeros to A. Then*

$$\sigma_i(A) = \sigma_i(B) = \sigma_i(C) \qquad i = 1, 2, \ldots, n.$$

For $k \in \{2, \ldots, n-1\}$ it is convenient to denote by $A_{[k]}$ the k-by-k submatrix of A obtained by deleting its *last* $n-k$ rows and columns; $A_{\langle k \rangle} \in M_{n-k+1}$ denotes the submatrix obtained by deleting the *first* $k-1$ rows and columns. Notice that the entry a_{kk} appears in both $A_{[k]}$ and $A_{\langle k \rangle}$.

We can now prove the two theorems. Since each is known already for $k = 1$ and $k = n$, the proofs will be given for an arbitrary $k \in \{2, \ldots, n-1\}$.

Proof of Theorem 1. The Schur triangularization theorem permits us to write $A = UTU^*$, where $U \in M_n$ is unitary and $T = [t_{ij}] \in M_n$ is upper triangular with $t_{ii} = \lambda_i(A) = \lambda_i(T)$. Moreover, $\sigma_i(T) = \sigma_i(A)$ by unitary invariance and Lemma 3 guarantees that $\sigma_i(T) \geqslant \sigma_i(T_{[k]})$. Thus,

$$\begin{aligned} \sigma_1(A) \cdots \sigma_k(A) &= \sigma_1(T) \cdots \sigma_k(T) \geqslant \sigma_1(T_{[k]}) \cdots \sigma_k(T_{[k]}) = |\det T_{[k]}| \\ &= |\lambda_1(T_{[k]}) \cdots \lambda_k(T_{[k]})| = |\lambda_1(T) \cdots \lambda_k(T)| \\ &= |\lambda_1(A) \cdots \lambda_k(A)|. \end{aligned}$$

Proof of Theorem 2. Write $A = UTU^*$ as before, so that $A^m = UT^mU^*$. The upper triangular structure of T ensures that $(T_{[k]})^m = (T^m)_{[k]}$ and $(T_{\langle k \rangle})^m = (T)^m_{\langle k \rangle}$. Thus, the same reasoning used in the proof of Theorem 1 gives the lower bound

$$\sigma_k(A^m) = \sigma_k(T^m) \geqslant \sigma_k\big((T^m)_{[k]}\big) = \sigma_k\Big(\big(T_{[k]}\big)^m\Big).$$

To obtain an upper bound, we use the second inequality in (5) (with $i = 1$ and $p = k-1$) and Lemma 4 to write

$$\sigma_k(A^m) = \sigma_k(T^m) = \sigma_{1+(k-1)}(T^m) \leqslant \sigma_1\big((T^m)_{\langle k \rangle}\big) = \sigma_1\Big(\big(T_{\langle k \rangle}\big)^m\Big).$$

Thus, we have the two-sided bounds

$$\Big[\sigma_k\Big(\big(T_{[k]}\big)^m\Big)\Big]^{1/m} \leqslant \big[\sigma_k(A^m)\big]^{1/m} \leqslant \Big[\sigma_1\Big(\big(T_{\langle k \rangle}\big)^m\Big)\Big]^{1/m}.$$

By the two cases of Theorem 2 we have already established (limits involving the smallest and largest singular values of powers of a square matrix), we know that

$$\lim_{m \to \infty} \Big[\sigma_k\Big(\big(T_{[k]}\big)^m\Big)\Big]^{1/m} = |\lambda_k(T_{[k]})|$$

and

$$\lim_{m\to\infty}\left[\sigma_1\left(\left(T_{\langle k\rangle}\right)^m\right)\right]^{1/m} = |\lambda_1(T_{\langle k\rangle})|.$$

But since $\lambda_k(T_{[k]}) = \lambda_k(T) = \lambda_k(A)$ and $\lambda_1(T_{\langle k\rangle}) = \lambda_k(T) = \lambda_k(A)$, we are done.

For more information on these and similar inequalities see Chapter 2 of [**1**], Chapter 3 of [**4**] and Sections (7.3–4) of [**3**].

Acknowledgement. I am grateful to Prof. R. Horn for his encouragement and assistance in the preparation of this paper.

REFERENCES

1. I. C. Gohberg and M. G. Krein, Introduction to the Theory of Linear Nonselfadjoint Operators, Translations of Mathematical Monographs, Vol. 18, American Mathematical Society, Providence, 1969.
2. A. Horn, On the singular values of a product of completely continuous operators, *Proc. Nat. Acad. Sci.*, 36 (1950) 374–375.
3. R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, New York, 1985.
4. R. A. Horn and C. R. Johnson, Topics in Matrix Analysis, Cambridge University Press, New York, 1989.
5. H. Weyl, Inequalities between the two kinds of eigenvalues of a linear transformation, *Proc. Nat. Acad. Sci.*, 35 (1949) 408–411.
6. T. Yamamoto, On the extreme values of the roots of matrices, *J. Math. Soc. Japan*, 19 (1967) 173–178.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

A Decomposition Theorem for Finite Automata

TERESA L. SANTA COLOMA AND RALPH P. TUCCI
Department of Mathematical Sciences, Loyola University, New Orleans, LA 70118

Automata theory is an important branch of theoretical computer science, and the Krohn-Rhodes theorem [**1**] is one of the best known theorems in this area. It states that the semigroup of a finite automaton can be covered by a wreath product of semigroups which are either finite simple groups or semigroups of the form U^1, where U is a right zero semigroup containing two elements and U^1 is the semigroup consisting of U with identity adjoined. This type of decomposition is useful, for example, when examining decompositions associated with VLSI (see [**4**]). Unfortunately, undergraduate students and even beginning graduate students usually do not have either the mathematical background or the mathematical maturity to understand a proof of this theorem. In this paper we present a decomposition theorem that is easier to understand than the Krohn-Rhodes theorem and could serve to give a student the flavor of the Krohn-Rhodes theorem without requiring an extensive technical background. For proofs of the Krohn-Rhodes theorem, the interested reader is referred to [**1**] and [**3**]. We assume the reader is familiar with the basic theory of semigroups as presented in [**2**].

DEFINITION. A *finite state automaton* (or *machine*) M is an ordered triple $M = (Q, \Sigma, F)$ where Q is a finite set called the set of *states*, Σ is a finite set called the *input alphabet*, and F is a *partial function* F: $Q \times \Sigma \to Q$.

We think of the machine as reading a string of symbols, one at a time, from the input alphabet Σ; when the machine is in state q reading symbol σ, then the machine goes to state $F(q, \sigma)$. Each element $\sigma \in \Sigma$ induces a partial function F_σ: $Q \to Q$ defined by $F_\sigma(q) = F(q, \sigma)$. Sometimes we will abbreviate $F(q, \sigma)$ by $q\sigma$.

Automata are often represented by directed graphs with labels on the edges. For example, suppose that the flowchart in FIGURE 1(a) represents a program that branches to statement A or to statement B depending on whether its input is a or b. Then the automaton in FIGURE 1(b) models this behavior.

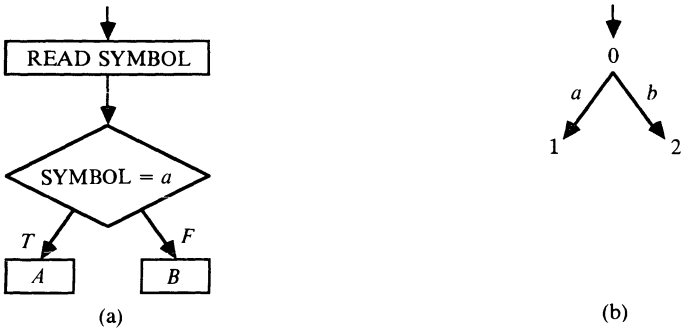
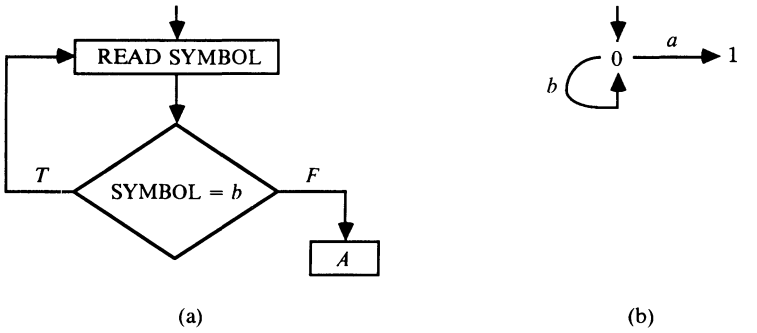


FIG. 1.

As another example, suppose the flowchart in FIGURE 2(a) represents a program which keeps reading b's until it reads an a, at which point it executes statement A. The automaton in FIGURE 2(b) models this behavior.



(a) (b)

FIG. 2.

Let Σ^* denote the set of all finite strings over Σ. Then we can extend the definition of F from elements of Σ to strings over Σ in the obvious manner; i.e., if $w \in \Sigma^*$, and $w = \sigma w'$ for some $\sigma \in \Sigma$ and $w' \in \Sigma^*$, then $F(q, w) = F(F(q, \sigma), w')$ for all $q \in Q$. As before, each string s over Σ^* induces a partial function F_s: $Q \to Q$ defined by $F_s(q) = F(q, s)$.

DEFINITION. Let $M = (Q, \Sigma, F)$ be an automaton, and take Σ^* to be a semigroup under concatenation. Define the congruence $\equiv$ on Σ^* by saying that if s, $s' \in \Sigma^*$, then $s \equiv s'$ if and only if $qs = qs'$ for all $q \in Q$. The *semigroup of the machine* M, denoted $S(M)$ or simply S, is the semigroup $\Sigma^*/\equiv$.

Note that S is a semigroup with identity, where the identity of S is induced by the empty string. Also note that we could define the congruence $\equiv$ by saying that $s \equiv s'$ if and only if s and s' induce the same partial function on Q; i.e., $s \equiv s'$ if and only if $F_s = F_{s'}$. Further, if s, $s' \in \Sigma^*$ and if the partial function $F_s \circ F_{s'}$ induced by the string ss' is undefined, then we adjoin a zero element 0 to S and define $ss' = 0$. (We can think of 0 as inducing a partial function on Q with domain ϕ.)

In this paper we decompose the semigroup of an automaton into simpler semigroups according to the next two definitions.

DEFINITION. Let $M_1 = (Q, \Sigma_1, F_1)$ and $M_2 = (Q, \Sigma_2, F_2)$ be two automata with the same state set. Then M_2 *covers* M_1, denoted $M_1 \leqslant M_2$, if there is a 1–1 function g: $\Sigma_1 \to \Sigma_2$ such that, for every $\sigma_1 \in \Sigma_1$ and $q \in Q$, either $F_1(q, \sigma_1)$ is undefined or $F_1(q, \sigma_1) = F_2(q, \sigma_2)$ where $\sigma_2 = g(\sigma_1)$. The function g is said to be a *covering of* M_1 *by* M_2. (The reader should note that this definition of covering is not standard; see [3, p. 43].)

This situation is best described by the commutative diagram in FIGURE 3, where $g(\sigma_1) = \sigma_2$ and where I denotes the identity function on Q. The expression (I, g) denotes the function from $Q \times \Sigma_1$ to $Q \times \Sigma_2$ defined by $(I, g)(q, \sigma_1) = (I(q), g(\sigma_1)) = (q, \sigma_2)$ for all $(q, \sigma_1) \in Q \times \Sigma_1$.

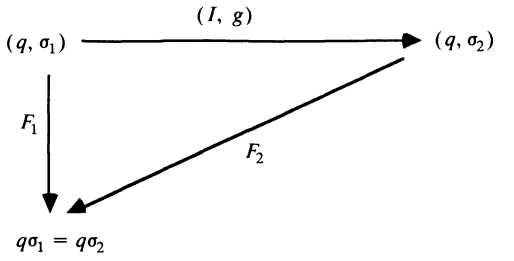


FIG. 3.

Intuitively, we can say that M_2 covers M_1 if M_2 can do everything that M_1 does. For example, in FIGURE 4 M_2 covers M_1; the covering g is the function from the alphabet of M_1 to the alphabet of M_2 given by $g(a) = c$ and $g(b) = d$. Note that $F_1(0, b)$ is undefined in M_1, while $F_2(0, g(b)) = F_2(0, d) = 1$ is defined in M_2.



FIG. 4.

DEFINITION. Let $M_1 = (Q, \Sigma_1, F_1)$ and $M_2 = (Q, \Sigma_2, F_2)$ be two automata with the same state set and disjoint alphabets. The *alphabet direct sum of* M_1 *and* M_2, denoted $M_1 \oplus M_2$, is the machine $M = (Q, \Sigma_1 \cup \Sigma_2, F)$ with $F(q, \sigma) = F_j(q, \sigma)$ for all $q \in Q$, where $j = 1$ or $j = 2$ depending on whether $\sigma \in \Sigma_1$ or $\sigma \in \Sigma_2$.

We are now ready to state the decomposition theorem. This theorem shows how we can cover the semigroup of a machine by a group with zero and a semigroup with no units except for the identity.

THEOREM. *Let S be the semigroup of an automaton M. Then there are automata M_1 and M_2 such that $M \leqslant M_1 \oplus M_2$ and (1) $S(M_1)$ is a group with zero which is a homomorphic image of $S(M)$; (2) $S(M_2)$ is made up of a subsemigroup of $S(M)$ which contains no units, together with an identity element.*

Proof. Let $S = S(M)$, and let $J = \{s \in S | F_S \text{ is not an onto function}\}$. Then J is a two-sided ideal of S (which therefore contains no units) and S/J is a group with zero. Take $M_1 = (Q, S/J, F_1)$; for each $q \in Q$ and $sJ \in S/J$, $s \notin J$, let $F_1(q, sJ) = F(q, s)$, and let $F(q, J)$ be undefined. Take $M_2 = (Q, J, F_2)$; for each $q \in Q$ and $s \in J$ let $F_2(q, s) = F(q, s)$. Define the function g: $S \to (S/J) \cup J$ by $g(s) = sJ \in S/J$ if $s \notin J$, and $g(s) = s \in J$ if $s \in J$. It is now easy to see that $M_1 \oplus M_2$ covers M. Let $s \in S$; if $s \in J$ then $F(q, s) = F_2(q, s)$ and if $s \notin J$ then $F(q, s) = F_1(q, sJ)$. The other statements in the theorem follow immediately.

Example. Let M be the machine in FIGURE 5. The semigroup of M is $S = \{1, a, b, ab, ba, aba, 0\}$ where $a^2 = 1$, $b^2 = b$, and $bab = 0$. The ideal J as defined

in the theorem is $J = \langle b \rangle = \{b, ab, ba, aba, 0\}$. The machine M_1 in FIGURE 6 has semigroup $S(M_1) = \{1, a, 0\} \cong Z_2 \cup \{0\}$ which is a group with zero, the machine M_2 has semigroup $S(M_2) = J^1$ (the semigroup J with identity adjoined) where J contains no units, and $M_1 \oplus M_2$ covers M. Note that the semigroups of M_1 and M_2 are simpler than the semigroup of M, even though the diagrams are a bit more complicated.

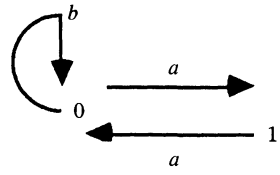


FIG. 5.



FIG. 6.

Acknowledgments. The authors would like to thank the referees for their helpful suggestions which greatly improved this paper.

REFERENCES

1. M. A. Arbib, The Algebraic Theory of Machines, Languages, and Semigroups, Academic Press, New York, 1968.
2. A. H. Clifford and G. B. Preston, The Algebraic Theory of Semigroups, vol. 1, American Math. Society, Providence, R.I., 1961.
3. W. M. L. Holcombe, Algebraic Automata Theory, Cambridge University Press, New York, 1982.
4. Y. Hou, Trinity algebra and its application to machine decomposition, *Information Processing Letters*, 26 (1987/88) 127–134.

A Counterexample in Galois Theory

THOMAS W. HUNGERFORD
Department of Mathematics, Cleveland State University, Cleveland, OH 44115

A typical first-year modern algebra course culminates in the proof of Galois' solvability criterion:

THEOREM. *Let F be a field of characteristic 0 and $f(x) \in F[x]$. If the equation $f(x) = 0$ is solvable by radicals, then the Galois group of $f(x)$ is a solvable group.**

*The converse is also true, but its proof is often omitted because this is all that's needed to demonstrate the insolvability of the quintic.

The proof involves a subtle point that is often missed by students and, alas, by the authors of several current modern algebra texts. In this note we present a simple example that shows why a certain technical lemma is necessary for a valid proof.

To understand the difficulty, we first review the usual proof. By the definition of "solvable by radicals," the splitting field K of $f(x)$ is contained in a radical extension L of F, that is, there is a tower of fields

$$F \subseteq F(u_1) \subseteq F(u_1, u_2) \subseteq \cdots \subseteq F(u_1, u_2, \ldots, u_t) = L, \qquad (*)$$

such that $K \subseteq L$,

$$u_1^{n_1} \in F \text{ and } u_i^{n_i} \in F(u_1, \ldots, u_{i-1}) \quad \text{for each } i \geqslant 1.$$

The next step is to adjoin all the n_ith roots of unity. They can be adjoined one at a time, or in a single step by letting n be the least common multiple of $n_1, \ldots, n_t$ and adjoining a primitive nth root of unity β. The powers of β include all the n_ith roots of unity. Since $\beta^n = 1 \in F$, the tower

$$F \subseteq F(\beta) \subseteq F(\beta, u_1) \subseteq F(\beta, u_1, u_2) \subseteq \cdots \subseteq F(\beta, u_1, u_2, \ldots, u_t) = E \quad (**)$$

is also a radical extension of F. Furthermore,

1. $F(\beta)$ a normal extension of F (it is the splitting field of $x^n - 1$). Its Galois group is abelian.**

2. Each $F(\beta, u_1, \ldots, u_i)$ is a normal extension of $F(\beta, u_1, \ldots, u_{i-1})$ (it is the splitting field of $x^{n_i} - u_i^{n_i}$). Its Galois group is abelian.**

Under the Galois correspondence, the sequence of fields (**) corresponds to a sequence of groups

$$G \supseteq G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_t = (1),$$

in which G is the Galois group of E over F and each G_i is normal in the preceding group. The Fundamental Theorem shows that the quotient G/G_0 is isomorphic to the Galois group of $F(\beta)$ over F (and hence is abelian by 1) and each quotient G_i/G_{i-1} is isomorphic to the Galois group of $F(\beta, u_1, \ldots, u_i)$ over $F(\beta, u_1, \ldots, u_{i-1})$ (and hence is abelian by 2). Therefore G is solvable. The proof is completed by showing that the Galois group of K over F (the Galois group of $f(x)$) is a homomorphic image of G and hence solvable.***

And now the sticky point: for this proof to be valid, the field E in (**) must be a *normal* extension of F (otherwise the Fundamental Theorem and the Galois correspondence can't be used). Most students tacitly assume that the construction of E given above guarantees its normality over F. Unfortunately, this may *not* be the case, despite some implicit and explicit statements to the contrary in several texts.

To see why E need not be normal over F, consider the tower

$$\mathbf{Q} \subseteq \mathbf{Q}(\sqrt{2}\,i) \subseteq \mathbf{Q}\left(\sqrt{2}\,i, \sqrt[4]{2}\,(1 - i)\right) = L.$$

L is a radical extension of the field $\mathbf{Q}$ of rational numbers because

$$(\sqrt{2}\,i)^2 = -2 \in \mathbf{Q} \quad \text{and}$$

$$\left(\sqrt[4]{2}\,(1 - i)\right)^2 = -2\sqrt{2}\,i \in \mathbf{Q}(\sqrt{2}\,i).$$

**See [4] or [5] for a proof of 1 and [3], [4], or [5] for a proof of 2.

***See [3] for details.

This is just the tower (*) above with $F = \mathbf{Q}$, $t = 2$, $u_1 = \sqrt{2}\,i$, $u_2 = \sqrt[4]{2}\,(1 - i)$, $n_1 = 2 = n_2$, and $n = \operatorname{lcm}(2, 2) = 2$. To obtain the tower (**) we must adjoin a primitive 2nd root of unity. But in this case $\mathbf{Q}$ already contains the primitive 2nd root $\beta = -1$. So $\mathbf{Q} = \mathbf{Q}(\beta)$ and the tower (**) coincides with (*):

$$\mathbf{Q} \subseteq \mathbf{Q}(\sqrt{2}\,i) \subseteq \mathbf{Q}\left(\sqrt{2}\,i, \sqrt[4]{2}\,(1 - i)\right) = E.$$

Furthermore, *the dimension* $[E : \mathbf{Q}]$ *is at most 4* because $(\sqrt{2}\,i)^2 \in \mathbf{Q}$ and $(\sqrt[4]{2}\,(1 - i))^2 \in \mathbf{Q}(\sqrt{2}\,i)$.

We now show that E is *not* normal over $\mathbf{Q}$. You can readily verify that $\pm \sqrt[4]{2}\,(1 - i)$ and $\pm \sqrt[4]{2}\,(1 + i)$ are the roots of $p(x) = x^4 + 8$, from which it follows that $p(x)$ is irreducible in $\mathbf{Q}[x]$. E contains the root $u = \sqrt[4]{2}\,(1 - i)$ of $p(x)$. If E were normal over $\mathbf{Q}$, $p(x)$ would split in $E[x]$ and the root $v = \sqrt[4]{2}\,(1 + i)$ would be in E. Hence E would contain both $v/u = i$ and $(v - u)/2i = \sqrt[4]{2}$ and we would have

$$\mathbf{Q} \subseteq \mathbf{Q}\left(\sqrt[4]{2}\right) \subseteq \mathbf{Q}\left(\sqrt[4]{2}, i\right) \subseteq E.$$

But $x^4 - 2$ is the minimal polynomial of $\sqrt[4]{2}$ over $\mathbf{Q}$, and $x^2 + 1$ is the minimal polynomial of i over $\mathbf{Q}(\sqrt[4]{2})$ (all of whose elements are real). Consequently,

$$[E : \mathbf{Q}] \geqslant \left[\mathbf{Q}\left(\sqrt[4]{2}, i\right) : \mathbf{Q}\right] = \left[\mathbf{Q}\left(\sqrt[4]{2}, i\right) : \mathbf{Q}\left(\sqrt[4]{2}\right)\right]\left[\mathbf{Q}\left(\sqrt[4]{2}\right) : \mathbf{Q}\right] = 2 \cdot 4 = 8.$$

This contradicts the fact that $[E : \mathbf{Q}] \leqslant 4$. Therefore E is not normal over $\mathbf{Q}$.

Once noted, the sticky point is easily taken care of by starting the proof of the Theorem with this!

LEMMA. *If F has a radical extension, then F has a normal radical extension.*

Then one can assume that L is normal over F in (*), from which it follows easily that E is normal over F in (**), so that the proof outlined above is valid.

The Lemma (or an equivalent) appears in the graduate texts [**4**], [**5**] and [**6**], as well as in undergraduate texts, such as [**1**] and [**3**] (where the proof is a starred exercise). The problem is avoided in [**2**] and [**7**] by restricting the discussion to normal radical extensions.

But in modern algebra texts published since 1986 (and a few earlier ones), the reader should be careful. Even when the Lemma is stated, its proof may be faulty. Most of these books present either a fallacious proof of the Theorem or Lemma (counterexample above) or a suspiciously incomplete proof.

The incomplete arguments typically say "adjoin all necessary roots of unity," but give no proof that this guarantees the normality of E over F. In this case, depending on the interpretation of "necessary," the example above may not be a counterexample (if you adjoin the 4th root of unity i between $\mathbf{Q}$ and $\mathbf{Q}(\sqrt{2}\,i)$, the resulting top field in (**) is $\mathbf{Q}(i, \sqrt[4]{2})$, which is normal over $\mathbf{Q}$).

REFERENCES

1. G. Birkhoff and S. Mac Lane, A Survey of Modern Algebra, 4th Edition, Macmillan, N.Y., 1977.
2. J. Fraleigh, A First Course in Abstract Algebra, 3rd Edition, Addison-Wesley, Reading, Mass., 1982.
3. I. N. Herstein, Topics in Algebra, 2nd Edition, Xerox, Lexington, Mass., 1975.
4. T. W. Hungerford, Algebra, Springer-Verlag, New York-Berlin-Heidelberg, 1974.
5. N. Jacobson, Basic Algebra I, 2nd Edition, Freeman, San Francisco, 1985.
6. S. Lang, Algebra, 2nd Edition, Addison-Wesley, Menlo Park, Calif., 1984.
7. _____, Undergraduate Algebra, Springer-Verlag, New York-Berlin-Heidelberg, 1987.

A Simple Algorithm for Finding Eigenvalues and Eigenvectors for 2×2 Matrices

TYRE A. NEWTON
Department of Mathematics, Washington State University, Pullman, WA 99164

The solution of the 2×2 matrix eigenvalue problem is of considerable significance in linear algebra, and is basic in linear systems theory. Moreover, this problem can arise whenever you have a nonlinear 2nd order system, whether it is composed of algebraic, differential, or difference equations. A technique that is commonly used in the nonlinear case is presented in [**1**, Chap. 9] where we find at each critical point of the system an associated matrix with real elements, namely,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{1}$$

For each such matrix we determine its eigenvalues and corresponding eigenvectors.

The traditional procedure for solving this eigenvalue problem is to first solve the quadratic characteristic equation of A for the eigenvalues and then for each eigenvalue solve a homogeneous system of linear equations to get the corresponding space of eigenvectors. In the method presented here, we solve a different quadratic equation related to A, and from its roots determine immediately both the eigenvalues and corresponding eigenvectors.

To facilitate this, we say that if λ is an eigenvalue and $\mathbf{v}$ is a corresponding eigenvector of A, then the pair $[\lambda, \mathbf{v}]$ is an *eigenpaar* of A. Since any nonzero multiple of an eigenvector is an eigenvector corresponding to the same eigenvalue, we agree that $[\lambda, \mathbf{v}] = [\lambda, a\mathbf{v}]$ for all real $a \neq 0$. Also associate with the matrix A given in (1) the equation

$$bm^2 + (a - d)m - c = 0. \tag{2}$$

It simplifies things and there is no inconsistency to state that ∞ is a root of (2) whenever $b = 0$. The multiplicity of ∞ as a root of (2) is 1 if $b = 0$ and $a - d \neq 0$, and 2 if $b = 0$ and $a - d = 0$, but $c \neq 0$.

Our procedure is given in the following.

PROPOSITION. *Given the matrix A in* (1).
(a) *The pair*

$$\left[a + bm, \begin{pmatrix} 1 \\ m \end{pmatrix}\right] \tag{3}$$

is an eigenpaar of A if and only if m is a finite root of (2).

(b) *The pair*

$$\left[d, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right] \tag{4}$$

is an eigenpaar of A if and only if ∞ *is a root of* (2).

(c) *If at least one of the coefficients in* (2) *is nonzero, then the space of eigenvectors for each eigenvalue found above is one-dimensional. If all three coefficients are zero, then*

$$\left[a, \begin{pmatrix} x \\ y \end{pmatrix}\right] \tag{5}$$

is an eigenpaar of A for any x and y that are not both zero.

Proof. For ease of notation we will define the vectors

$$\mathbf{v}_m = \begin{pmatrix} 1 \\ m \end{pmatrix} \quad \text{and} \quad \mathbf{v}_\infty = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

If m is real, then m is the slope of $\mathbf{v}_m$ while the slope of $\mathbf{v}_\infty$ is infinite.

For part (a), assume that m is a finite root of (2). When we rewrite (2) as

$$m(a + bm) = c + dm, \tag{6}$$

we conclude that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 \\ m \end{pmatrix} = (a + bm)\begin{pmatrix} 1 \\ m \end{pmatrix}. \tag{7}$$

Consequently (3) is an eigenpaar of A. Conversely, if (3) is an eigenpaar of A for finite m, then (7) is true from which (6) follows and m is a root of (2).

For part (b), consider the equation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = d\begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

We see that (4) is an eigenpaar of A if and only if $b = 0$ in which case ∞ is a root of (2).

For part (c), first assume that (2) has at least one nonzero coefficient so that it has at most two roots. When we consider the equations

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 \\ m \end{pmatrix} = \lambda\begin{pmatrix} 1 \\ m \end{pmatrix},\ b \neq 0 \quad \text{and} \quad \begin{pmatrix} a & 0 \\ c & d \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \lambda\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

we see that if $\mathbf{v}_m$ or $\mathbf{v}_\infty$ is an eigenvector, then necessarily the corresponding eigenvalue is $a + bm$ or d respectively. Since any nonzero vector in R^2 can be written uniquely as a constant multiple either of $\mathbf{v}_m$ for some real m, or of $\mathbf{v}_\infty$, then any eigenpaar defined by (3) or (4) defines a one-dimensional space of eigenvectors. However, if all three coefficients of (2) are zero, then ∞ and any finite m will be a root. In this case A is diagonal with equal diagonal elements and it follows from (3) and (4) that $[a, \mathbf{v}_m]$ for any m and $[a, \mathbf{v}_\infty]$ are eigenpaars. That this is the same as (5) for any x and y now follows from our statement of equality for eigenpaars. □

Comments. In making phase plane plots of the linear system

$$d\mathbf{X}/dt = A\mathbf{X} \tag{8}$$

where A is given by (1), we find that m is a solution of (2) if and only if $y = mx$ contains linear trajectories of (8). These lines partition the phase plane up into

sectors which bound the phase trajectories of (8). Consequently, we are led to call (2) the *sector equation* of A, and its roots the *sector roots* of **A**.

A property common to the characteristic equation of A, namely

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0,$$

and the sector equation of A is that they have the same discriminant, namely

$$\Delta = (a - d)^2 + 4bc.$$

It is immediate that $\lambda = a + bm$ is the transformation that relates the characteristic equation and the sector equation.

The technique presented here facilitates the writing of a simple microcomputer program for finding all eigenpaars for a given 2×2 real matrix.

We conclude with some examples.

Example 1. $\Delta > 0$, $b = 0$. Given

$$A = \begin{pmatrix} 2 & 0 \\ 6 & 1 \end{pmatrix},$$

the sector equation is $m - 6 = 0$, so that the sector roots are ∞ and 6. Since for $m = 6$, $a + bm = 2$, and the eigenpaars are

$$\left[1, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right] \quad \text{and} \quad \left[2, \begin{pmatrix} 1 \\ 6 \end{pmatrix}\right].$$

Example 2. $\Delta < 0$. Given

$$A = \begin{pmatrix} 1 & 1 \\ -2 & 3 \end{pmatrix},$$

the sector equation is $m^2 - 2m + 2 = 0$, so that $m = 1 \pm i$, and $a + bm = 1 + (1 \pm i) = 2 \pm i$. Consequently the eigenpaars are

$$\left[2 \pm i, \begin{pmatrix} 1 \\ 1 \pm i \end{pmatrix}\right].$$

Example 3. $\Delta > 0$, $b \neq 0$. Given

$$A = \begin{pmatrix} 1 & -2 \\ -1 & 2 \end{pmatrix},$$

the sector equation is $-2m^2 - m + 1 = 0$ so that the sector roots are $m = 1/2, -1$. At $m = 1/2$, $a + bm = 0$. At $m = -1$, $a + bm = 3$ so that the eigenpaars are

$$\left[0, \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}\right] = \left[0, \begin{pmatrix} 2 \\ 1 \end{pmatrix}\right] \quad \text{and} \quad \left[3, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right]$$

Note that A is singular here but is nonsingular in Example 1.

Example 4. Infinity is a sector root of multiplicity two. Given

$$A = \begin{pmatrix} a & 0 \\ c & a \end{pmatrix}, \qquad c \neq 0,$$

the sector equation is $0m^2 + 0m + c = 0$ having ∞ as a double root. Then the only eigenpaar is

$$\left[a, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right].$$

REFERENCE

1. William E. Boyce, and Richard C. DiPrima, Elementary Differential Equations and Boundary Value Problems, 4th ed., Wiley, New York, 1986.

Magic Squares and Linear Algebra

ARNO VAN DEN ESSEN
Department of Mathematics, Catholic University of Nijmegen, Nijmegen, The Netherlands

Magic squares have been the subject of an enormous number of papers. The aim of this note is to show how magic squares can be used as examples in a first year course in linear algebra.

Let $\mathbf{Q}$ be the field of rationals and n a positive integer greater than 1. An n-by-n matrix M with entries in $\mathbf{Q}$ is called *semimagic*, or a *semimagic* square, if there exists an element of $\mathbf{Q}$ denoted by $\sigma(M)$ such that the sum of each row and each column equals $\sigma(M)$. The matrix M is called *magic*, or a *magic square*, if furthermore the traces of both M and JM equal $\sigma(M)$ (J is the n-by-n matrix satisfying $J_{ij} = 0$ if $i + j \neq n + 1$ and $J_{ij} = 1$ if $i + j = n + 1$). The set of magic squares of order n (resp., the set of semi-magic squares of order n) we denote by $\mathrm{Mag}_n(\mathbf{Q})$ (resp., $\mathrm{SMag}_n(\mathbf{Q})$). If no confusion is possible we drop the index n. Finally I denotes the identity matrix.

As announced above we now describe four examples where magic squares can be used to illustrate several concepts of linear algebra.

1. The first thing to observe is that both sets $\mathrm{SMag}_n(\mathbf{Q})$ and $\mathrm{Mag}_n(\mathbf{Q})$ are $\mathbf{Q}$-vector spaces. Then questions like "compute the dimension of both spaces" and "find a $\mathbf{Q}$-basis for each of them" (for small n) supply nice exercises [2].

2. Obviously $\mathrm{Mag}_n(\mathbf{Q})$ is a linear subspace of $M_n(\mathbf{Q})$, the set of all n-by-n matrices with entries in $\mathbf{Q}$. The following proposition, which describes how the vector space of magic squares sits in the vector space of semi-magic squares, illustrates the notions of trace of a matrix and direct sum decomposition of vector spaces.

PROPOSITION 2.1. $\mathrm{SMag}(\mathbf{Q}) = \mathrm{Mag}(\mathbf{Q}) \oplus \mathbf{Q}I \oplus \mathbf{Q}J$.

Proof. Denote by E the n-by-n matrix with entries all equal to 1. Obviously E is magic. The crucial property of E is that $EM = \sigma(M)E = ME$ for all semi-magic squares M. In particular $E^2 = \sigma(E)Ee = nE$ and more generally $E^p = n^{p-1}E$ for all $p \geqslant 1$.

i) First we show

$$\mathrm{Mag}(\mathbf{Q}) = \mathrm{Mag}^0(\mathbf{Q}) \oplus \mathbf{Q}E \quad \text{and} \quad \mathrm{SMag}(\mathbf{Q}) = \mathrm{SMag}^0(\mathbf{Q}) \oplus \mathbf{Q}E. \qquad (1)$$

where $\mathrm{Mag}^0(\mathbf{Q})$ (resp., $\mathrm{SMag}^0(\mathbf{Q})$) denotes the set of magic (resp., semi-magic) matrices with trace zero; namely, if M is magic (resp., semi-magic) then $M - \frac{1}{n}\mathrm{Tr}(M)E$ has trace zero and is magic (resp., semi-magic). That both sums are direct follows by taking traces.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, FREDERICK W. LUTTMAN, MARVIN MARCUS, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed *problems should be sent to the* MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

An asterisk () after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions *of published problems should arrive before June 30, 1990 at the* MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3361. *Proposed by William P. Wardlaw, U.S. Naval Academy, Annapolis, MD.*

Prove or give a counterexample to the following conjecture concerning the Euler totient function: For every even positive integer n there is an odd positive integer k such that

$$\varphi(k) = \varphi(n).$$

E 3362. *Proposed by Bjorn Poonen, Winchester, MA.*

Given a positive integer n, let M denote the set of n by n matrices with complex entries. If $A \in M$, let A^* denote the conjugate transpose of A. Prove that for each positive integer k and each $B \in M$ there exists a unique $A \in M$ such that

$$A(A^*A)^k = B.$$

E 3363. *Proposed by N. J. Fine, Deerfield Beach, FL.*

A printer is given three copies of the numeral 4 and an unlimited supply of each of the six symbols

$$+, \cdot, -, \div, \sqrt{\ }, \lfloor\ \rfloor,$$

as well as an unlimited supply of left and right parentheses. (Here $\lfloor t \rfloor$ denotes the greatest integer not exceeding the real number t.) Show that he can compose an expression for each positive integer N. For example,

$$19 = \left\lfloor \sqrt{4 \div \left(\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{4}}}}}} - \lfloor \sqrt{\sqrt{4}} \rfloor\right)} \right\rfloor.$$

E 3364. *Proposed by H. W. Lenstra, University of California, Berkeley.*

Determine whether there exist positive integers a, b, n, m such that $a \neq b$, $n \geqslant 2$, $m \geqslant 2$, and

$$\underbrace{a^{a^{\cdot^{\cdot^{\cdot^{a}}}}}}_{n} = \underbrace{b^{b^{\cdot^{\cdot^{\cdot^{b}}}}}}_{m}.$$

E 3365. *Proposed by Dean S. Clark, University of Rhode Island, Kingston, RI.*

Let S denote the sum of the integer weights that come with a two-pan balance scale. Give a necessary and sufficient condition on the weights that assures that all integer-weighted objects in the range of 1 to S can be weighed exactly. Use of weights on both pans is permitted.

E 3366. *Proposed by the Editors, University of Illinois at Urbana-Champaign.*

For $n = 1, 2, 3, \ldots$ determine the subset of $(0, 1)$ on which

$$\left(\frac{d}{dx}\right)^n \{\log x \cdot \log(1 - x)\} < 0.$$

SOLUTIONS OF ELEMENTARY PROBLEMS

Breaking a Rock into Pieces

E 3094 [1985, 427]. *Proposed by Louis Funar, University of Craiova, Romania.*

Let D be a convex body in $\mathbb{R}^3$, and let $\lambda(D) = \sup_\pi$ area $(\pi \cap D)$ where π is a variable plane. Prove that D can be divided into four parts D_1, D_2, D_3, D_4 so that $\lambda(D_i) < \lambda(D)$, $i = 1, 2, 3, 4$. Is it possible to divide D into three parts with the same property?

Solution by C. A. Rogers, University College, London. It is possible to divide D into two parts with the same property. Choose an interior point $\mathcal{O}$ of D. Choose a sequence $l_1, l_2, \ldots$ of chords of D passing through $\mathcal{O}$, the chords being dense in D. We choose a sequence $\varepsilon_1, \varepsilon_2, \ldots$ of small positive numbers decreasing rapidly to zero. For each $n \geqslant 1$, let C_n denote the set of all points of D lying at a distance less than ε_n from l_n. By choosing $\varepsilon_1, \varepsilon_2, \ldots$ suitably we can ensure that $\lambda(C_n) < 2^{-n}\lambda(D)$, for each $n \geqslant 1$.

Take $D_1 = \bigcup_{n=1}^{\infty} C_n$ and $D_2 = D \setminus D_1$. Then

$$\lambda(D_1) \leqslant \sum_{n=1}^{\infty} \lambda(C_n) < \lambda(D).$$

It is also clear that

$$\text{area } (\pi \cap D_2) < \lambda(D) \tag{1}$$

for each plane π.

Suppose it were true that $\lambda(D_2) = \lambda(D)$. Then it would be possible to choose a sequence $\pi_1, \pi_2, \ldots$ of planes so that

$$\text{area}\,(\pi_n \cap D_2) \to \lambda(D) \quad \text{as} \quad n \to \infty.$$

Further, these planes could be chosen so that they converge to some plane π. By (1), we would have

$$\text{area}\,(\pi \cap D_2) < \lambda(D).$$

So we could choose a constant μ and an open set G in π with $\pi \cap D_2 \subset G$ and

$$\text{area}\, G < \mu < \lambda(D).$$

Since D_2 is compact, points of $\pi_n \cap D_2$ cannot approach points of $\pi \setminus G$ as n tends to infinity. This would ensure that

$$\text{area}\,(\pi_n \cap D_2) < \mu < \lambda(D)$$

for all sufficiently large n. Thus our supposition is false and hence $\lambda(D_2) < \lambda(D)$, as required.

Editorial comment. Although a partial solution was given by the proposer, no complete solution was received before the customary deadline. The above solution was prepared by Professor Rogers at the request of the editors.

The proposer may have intended to place further restrictions on the "parts," but the above solution certainly disposes of the problem as published.

A Subtle Inequality

E 3116 [1985, 666]. *Proposed by M. Laub, Jerusalem, Israel.*

(a) If x and y are non-negative real numbers, prove that $x^x + y^y \geqslant x^y + y^x$.

(b) If $\{x_1, x_2, \ldots, x_n\}$ is a finite sequence of non-negative real numbers and $\{y_1, y_2, \ldots, y_n\}$ is any permutation of this sequence, prove or disprove that

$$x_1^{x_1} + x_2^{x_2} + \cdots + x_n^{x_n} \geqslant x_1^{y_1} + x_2^{y_2} + \cdots + x_n^{y_n}.$$

Solution by Ishai Ilani, Jerusalem College of Technology, Israel. Both parts are true.

(a) Without loss of generality we may assume that $x < y$, and we shall prove that

$$y^y - y^x - (x^y - x^x) > 0.$$

We shall distinguish two cases:

Case (i), $y > 1$. In this case we shall use the identity

$$y^y - y^x - (x^y - x^x) = \int_x^y \left(y^t \log y - x^t \log x\right) dt$$

and, since $y > \max\{1, x\}$, we have

$$y^t \log y > x^t \log y > x^t \log x.$$

Hence,

$$y^y - y^x - (x^y - x^x) > 0.$$

Case (ii), $x < y \leqslant 1$. In this case we consider the function f defined by $f(t) = y^y - y^t - (t^y - t^t)$, where y is fixed and $0 \leqslant t \leqslant y$. Obviously $f(y) = 0$. We shall prove that $f'(t) < 0$ for all t such that $0 < t < y$, from which follows immediately that $f(x) > 0$.

Clearly

$$f'(t) = g(t) + h(t),$$

where

$$g(t) = t^t \log t - y^t \log y$$

and

$$h(t) = t^t - yt^{y-1}.$$

Since

$$g(t) = -\int_t^y \frac{d}{ds}(s^t \log s)\, ds = -\int_t^y s^{t-1}(1 + t \log s)\, ds,$$

the condition $t \leqslant s \leqslant y \leqslant 1$ implies that $1 + t \log s \geqslant 1 + t \log t \geqslant 1 - e^{-1} > 0$. Hence $g(t) < 0$. In order to see that $h(t) < 0$, we consider the function k defined on $(0, \infty)$ by $k(z) = t^t - zt^{z-1}$, where t is fixed; we wish to show that $k(y) < 0$. Now $0 < t < y \leqslant 1$ and $k(t) = 0 > t^t - 1 = k(1)$ and $k(z) \to t^t$ as $z \to \infty$.

If $k(y)$ were non-negative, then k would have a zero in $[y, 1)$ and a zero in $(1, \infty)$ in addition to the smaller zero at t. Thus if $k(y) \geqslant 0$, the function k would have at least three positive zeros and hence k' would have at least two positive zeros. But, since $k'(z) = -t^{z-1}(1 + z \log t)$, we see that actually k' has only one positive zero. Thus we must have $h(t) = k(y) < 0$.

(b) It is enough to prove the assertion for cycles, i.e., that

$$x_1^{x_1} + x_2^{x_2} + \cdots + x_{n-1}^{x_{n-1}} + x_n^{x_n} \geqslant x_1^{x_2} + x_2^{x_3} + \cdots + x_{n-1}^{x_n} + x_n^{x_1} \qquad (*)$$

for all positive integers n and all nonnegative real numbers $x_1, \ldots, x_n$. We proceed by induction on n. The assertion is true for $n = 1, 2$. So suppose $n > 2$ and $(*)$ is known to be true if n is replaced by $n - 1$. Then

$$x_1^{x_1} + x_2^{x_2} + \cdots + x_{n-1}^{x_{n-1}} + x_n^{x_n} - \left(x_1^{x_2} + x_2^{x_3} + \cdots + x_{n-1}^{x_n} + x_n^{x_1}\right) = A + B,$$

where

$$A = \left(x_1^{x_1} + x_2^{x_2} + \cdots + x_{n-2}^{x_{n-2}} + x_{n-1}^{x_{n-1}}\right) - \left(x_1^{x_2} + x_2^{x_3} + \cdots + x_{n-2}^{x_{n-1}} + x_{n-1}^{x_1}\right),$$

$$B = \left(x_n^{x_n} - x_n^{x_1}\right) - \left(x_{n-1}^{x_n} - x_{n-1}^{x_1}\right).$$

By induction $A \geqslant 0$. To prove $B \geqslant 0$ let us put for brevity

$$x_1 = x, \quad x_{n-1} = y, \quad x_n = z,$$

so that B may be written

$$B = \left(z^z - z^x\right) - \left(y^z - y^x\right).$$

Again there are two cases:

Case I, one of the x_i is greater than 1. Without loss of generality we may assume that $x_n = \max_{1 \leqslant i \leqslant n} x_i$, since the cycle may be rotated if necessary. Then

$$B = z^z - z^x - \left(y^z - y^x\right) = \int_x^z \left(z^t \log z - y^t \log y\right) dt \geqslant 0,$$

since $z \geqslant \max(1, x, y)$ and, accordingly,

$$z^t \log z \geqslant y^t \log z \geqslant y^t \log y.$$

Thus $B \geqslant 0$ in case I.

Case II, none of the x_i is greater than 1. Without loss of generality we may assume $x_n = \min_{1 \leqslant i \leqslant n} x_i$, since the cycle may be rotated if necessary. Thus, in the notation introduced above, $z = x_n \leqslant \min(x_1, x_{n-1}) = \min(x, y)$. We consider the function F defined by $F(t) = t^t - t^x - (y^t - y^x)$. Clearly

$$\begin{aligned} F'(t) &= (t^t \log t - y^t \log y) + (t^t - xt^{x-1}) \\ &= g(t) + h^*(t), \end{aligned}$$

where g is as in part (a) and $h^*(t) = t^t - xt^{x-1}$. As in part (a) we have $g(t) < 0$ for $0 < t < y$ and $h^*(t) < 0$ for $0 < t < x$, so that $F'(t) < 0$ for $0 < t < \min(x, y)$. But

$$F(\min(x, y)) = F(x) = F(y) = 0,$$

and thus $F(t) \geqslant 0$ for $0 \leqslant t \leqslant \min(x, y)$. Since $z \leqslant \min(x, y)$, we have $B = F(z) \geqslant 0$ in case II. Thus the induction is complete and assertion (b) is proved.

Editorial comment. As noted in Case (i) of the above solution, the assertion of (a) is almost immediate if $\max(x, y) \geqslant 1$. Thus if $x < y$ and $y \geqslant 1$, we have

$$y^y - y^x = y^x(y^{y-x} - 1) \geqslant x^x(y^{y-x} - 1) \geqslant x^x(x^{y-x} - 1) = x^y - x^x.$$

However, the first inequality fails when $x < y < 1$, since then $y^{y-x} - 1$ is negative. Thus a more subtle argument seems to be required in case $\max(x, y) < 1$. Many would-be solvers either ignored this case or asserted incorrectly that the preceding trivial argument applies.

Part (a) was solved also by B. Brunson. Part (b) was partially solved by E. Rips and the proposer jointly.

A Polynomial Identity

E 3207 [1987, 456]. *Proposed by Geng-Zhe Chang, University of Science and Technology of China, Hefei, Anhui, China.*

If m is a positive integer, let

$$F_m(x) = \sum_{k=0}^{m} \left(1 - \frac{k}{m}\right)\binom{2m}{k} x^k (1 - x)^{2m-k}.$$

Show that

$$F_{m-1}(x) - F_m(x) = \frac{1}{2(2m-1)}\binom{2m}{m} x^m (1 - x)^m$$

for $m > 1$.

Solution I by the University of South Alabama Problem Group, Mobile. For $m > 1$, set $G_m(x) = F_{m-1}(x) - F_m(x)$. Using Leibniz's rule for the differentiation of a

product,

$$F_m^{(i)}(0) = \frac{(-1)^i(2m)!}{(2m-i)!} \sum_{k=0}^{\min\{i,m\}} \left(1 - \frac{k}{m}\right)\binom{i}{k}(-1)^k.$$

Hence $F_m(0) = 1$, $F_m'(0) = -2$, and $F_m^{(i)}(0) = 0$ for $2 \leqslant i \leqslant m$. This yields $G_m^{(i)}(0) = 0$ for $0 \leqslant i \leqslant m-1$. Also $(1-x)^{m+1}$ is a factor of $F_m(x)$. Thus $G_m(x) = a_m x^m(1-x)^m$ for some constant a_m, since the degree of G_m is at most $2m$. If $H_m(x) = G_m(x)/(1-x)^m$, then $a_m = H_m(1)$ is the coefficient of the last nonzero term in the sum defining $F_{m-1}(x)$, so that

$$a_m = H_m(1) = \left(1 - \frac{m-2}{m-1}\right)\binom{2m-2}{m-2} = \frac{1}{2(2m-1)}\binom{2m}{m}.$$

Solution II by students of the 1987 Mathematical Olympiad Program, U.S. Military Academy, West Point, NY. Let

$$B(t; n, x) = \sum_{j=0}^{t} \binom{n}{j} x^j(1-x)^{n-j}.$$

We claim that the polynomial identity

$$B(t; n, x) - B(t+1; n+2, x)$$
$$= \binom{n}{t} x^{t+2}(1-x)^{n-t} - \binom{n}{t+1} x^{t+1}(1-x)^{n-t+1} \qquad (*)$$

holds for all nonnegative integers t. To see this, suppose that $0 \leqslant x \leqslant 1$, and view the identity in the language of probability theory, specifically in terms of $n+2$ Bernoulli trials with success probability x. The identity $(*)$ may be rewritten in the equivalent form

$$B(t; n, x) + \binom{n}{t+1} x^{t+1}(1-x)^{n-t+1}$$
$$= B(t+1; n+2, x) + \binom{n}{t} x^{t+2}(1-x)^{n-t}. \qquad (**)$$

Now $B(t; n, x)$ is the probability of at most t successes in the first n trials regardless of what happens in trials $n+1$ and $n+2$, the second term on the left of $(**)$ is the probability of exactly $t+1$ successes in the first n trials followed by failures on trials $n+1$ and $n+2$, $B(t+1; n+2, x)$ is the probability of at most $t+1$ successes in the $n+2$ trials, and the second term on the right of $(**)$ is the probability of exactly t successes in the first n trials followed by successes on trials $n+1$ and $n+2$. Using these interpretations, we see that both sides of $(**)$ give the probability of the following event: There are at most $t+2$ successes in $n+2$ Bernoulli trials with the restriction that $t+2$ successes can occur only in the case of exactly t successes in the first n trials followed by two successes in the last two trials. Thus $(**)$ holds for $0 \leqslant x \leqslant 1$ and hence $(**)$ and $(*)$ hold for all values of x.

Now, note that $F_m(x) = B(m-1; 2m, x) - 2xB(m-2; 2m-1, x)$ for $m > 1$. Using the identity (*), we find

$$\begin{aligned} F_{m-1}(x) - F_m(x) &= B(m-2; 2m-2, x) - B(m-1; 2m, x) \\ &\quad - 2x\{B(m-3; 2m-3, x) - B(m-2; 2m-1, x)\} \\ &= \binom{2m-2}{m-2} x^m (1-x)^m - \binom{2m-2}{m-1} x^{m-1}(1-x)^{m+1} \\ &\quad - 2x\left\{\binom{2m-3}{m-3} x^{m-1}(1-x)^m \right. \\ &\qquad \left. - \binom{2m-3}{m-2} x^{m-2}(1-x)^{m+1}\right\} \\ &= \left\{\binom{2m-2}{m-2} - 2\binom{2m-3}{m-3}\right\} x^m (1-x)^m \\ &= \frac{1}{2(2m-1)} \binom{2m}{m} x^m (1-x)^m \end{aligned}$$

for all $m > 2$. The assertion of the problem can be verified directly for $m = 2$.

Solution III and generalization by David G. Robinson, Emory University, Atlanta, GA. We define the symbol

$$\binom{n}{k}_{a,b}$$

recursively by putting

$$\binom{n}{0}_{a,b} = a \quad \text{for} \quad n \geqslant 1, \qquad \binom{n}{n}_{a,b} = b \quad \text{for} \quad n \geqslant 1,$$

$$\binom{n}{k}_{a,b} = \binom{n-1}{k}_{a,b} + \binom{n-1}{k-1}_{a,b} \quad \text{for} \quad 1 \leqslant k \leqslant n-1.$$

By induction we may express this symbol in terms of ordinary binomial coefficients as follows:

$$\binom{n}{k}_{a,b} = a\binom{n-1}{k} + b\binom{n-1}{k-1} = \left\{a\left(1 - \frac{k}{n}\right) + b\frac{k}{n}\right\}\binom{n}{k}$$

$$\text{for} \quad 1 \leqslant k \leqslant n-1.$$

In particular

$$\binom{n}{k}_{1,-1} = \left(1 - \frac{2k}{n}\right)\binom{n}{k} \quad \text{for} \quad 0 \leqslant k \leqslant n, \; n \geqslant 1.$$

In terms of generating functions we have

$$\sum_{k=0}^{n} \binom{n}{k}_{a,b} x^k y^{n-k} = (ay + bx)(y + x)^{n-1} \qquad (n \geqslant 1).$$

Defining

$$A_n(a, b; x, y) = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{k}_{a,b} x^k y^{n-k}, \qquad (n \geqslant 1)$$

we see that $F_m(x) = A_{2m}(1, -1; x, 1-x)$. The recurrence formula for the coefficients yields the following recurrence for the polynomials:

$$A_n(a, b; x, y) - (x+y)A_{n-1}(a, b; x, y) = (-1)^n \binom{n-1}{\lfloor n/2 \rfloor}_{a,b} x^{\lfloor (n+1)/2 \rfloor} y^{\lfloor n/2 \rfloor} \qquad (n > 1).$$

Applying this twice, we obtain (for $m > 1$)

$$A_{2m}(a, b; x, y) - (x+y)^2 A_{2m-2}(a, b; x, y) = \binom{2m-2}{m}_{a,b} x^m y^m - \binom{2m-2}{m-2}_{a,b} x^{m+1} y^{m-1},$$

which reduces to the desired recurrence for F when $a = 1$, $b = -1$, and $y = 1 - x$.

Solved also by G. E. Andrews, K. L. Bernstein, J. Fitch, T. Hermann (Hungary), Kanji Ikari & Shizuka Nakano & Akira Teramoto & Kaori Yagow & Hiroshi Zaizen jointly (Japan), E. Lee, O. P. Lossers (The Netherlands), Y. S. Sathe & R. G. Sheno jointly (India), J. H. Steelman, and the proposer.

Asymptotic Behavior of a Recurrence

E 3229 [1987, 787]. *Proposed by Gary E. Stevens, Hartwick College, Oneonta, New York.*

Determine the asymptotic behavior of the sequence defined by the recurrence

$$a_0 = 1, \qquad a_n = n^{1/a_{n-1}} \quad \text{for} \quad n > 0.$$

Solution by Preben Alsholm, Danmarks Ingeniorakademi, Lyngby, Denmark. We will show

$$a_{2k} = 2k - (\log 2k)^2 + O\left(\frac{(\log k)^4}{k}\right),$$

$$a_{2k+1} = 1 + \frac{\log(2k+1)}{2k+1} + O\left(\frac{(\log k)^3}{k^2}\right).$$

Let $b_{2k} = \log a_{2k}/\log 2k$ for $k \geqslant 1$. Applying the recurrence twice yields

$$b_{2k+2} = e^{-\log(2k+1)\exp(-b_{2k}\log 2k)}. \tag{1}$$

We will show that the sequence $\{b_{2k}\}$ is increasing for $k \geqslant k_0$, provided $k_0 \geqslant 2$ and

$$b_{2k_0} > \max\left\{ b_{2k_0-2}, \frac{(2k_0-2)^2}{(2k_0-3)(2k_0-1)\log(2k_0-1)} \right\}. \tag{2}$$

Using a calculator, one finds that $k_0 = 15$ satisfies (2). Suppose then by induction that $b_{2k} > b_{2k-2} \geqslant b_{2k_0-2}$ for some $k \geqslant k_0$. Using (1), we see that $b_{2k+2} > b_{2k}$ if and only if

$$e^{(b_{2k}-b_{2k-2})\log(2k-2)+b_{2k}\log\{2k/(2k-2)\}} > \frac{\log(2k+1)}{\log(2k-1)}.$$

Let L denote the left side of this desired inequality. By the induction hypothesis, we find

$$L > e^{b_{2k}\log\{2k/(2k-2)\}} \geqslant e^{b_{2k_0}\log\{2k/(2k-2)\}} \geqslant 1 + b_{2k_0}\log\frac{2k}{2k-2}$$

$$= 1 + b_{2k_0}\log\left(1 + \frac{1}{k-1}\right).$$

Since $\log(1+x) \geqslant x - x^2/2 \geqslant x(1 - x_0/2)$ for $0 \leqslant x \leqslant x_0$, we set $x_0 = 1/(k_0 - 1)$ to obtain

$$L > 1 + b_{2k_0}\frac{1}{k-1}\cdot\frac{2k_0 - 3}{2k_0 - 2}.$$

On the other hand, the inequality $\log(1+x) < x$ for positive x yields

$$\frac{\log(2k+1)}{\log(2k-1)} < 1 + \frac{2}{(2k-1)\log(2k-1)}$$

$$< 1 + \frac{2k_0 - 2}{(2k_0 - 1)\log(2k_0 - 1)}\cdot\frac{1}{k-1},$$

since $(a-1)/(a\log a)$ is a decreasing function of a for $a > 1$. Applying (2) to this inequality, we obtain

$$\frac{\log(2k+1)}{\log(2k-1)} < 1 + b_{2k_0}\cdot\frac{2k_0 - 3}{2k_0 - 2}\cdot\frac{1}{k-1} < L.$$

Thus, $\{b_{2k}\}$ is an increasing sequence for $k \geqslant k_0$. Since $a_n \geqslant 1$ for all $n \geqslant 0$ and thus $a_n = n^{1/a_{n-1}} \leqslant n$, it follows that $b_{2k} \leqslant 1$ for all $k \geqslant 1$, and hence $\{b_{2k}\}$ converges to some α in $(0,1]$. This implies b_{2k} exceeds $\alpha/2$ for sufficiently large k, at which point (1) yields

$$-\log b_{2k+2} = \frac{\log(2k+1)}{(2k)^{b_{2k}}} \leqslant \frac{\log(2k+1)}{(2k)^{\alpha/2}}.$$

Hence $\log b_{2k}$ converges to 0 and $\alpha = 1$. It follows immediately from $a_{2k-1} = b_{2k}^{-1}$ that a_{2k-1} decreases to 1 as $k \to \infty$ (for $k \geqslant k_0$). Since b_{2k} increases to 1, we find for any $\varepsilon > 0$ that $1 - \varepsilon < b_{2k} < 1$ for $k \geqslant k(\varepsilon)$. Thus $(2k)^{1-\varepsilon} < a_{2k} < 2k$ for $k \geqslant k(\varepsilon)$. By using

$$a_n = e^{(\log n)/a_{n-1}} \tag{3}$$

with $n = 2k+1$, we find

$$\exp\left(\frac{\log(2k+1)}{2k}\right) < a_{2k+1} < \exp\left(\frac{\log(2k+1)}{(2k)^{1-\varepsilon}}\right).$$

Since a_{2k+1} does not equal the preceding lower bound, we cannot get an expansion for a_{2k+1} solely by expanding the series for the exponential. However, since $\log x = o(x^{\delta})$ for all $\delta > 0$, we obtain the approximation

$$a_{2k+1} = 1 + o(k^{\varepsilon_1 - 1}) \quad \text{for} \quad \varepsilon_1 > \varepsilon. \tag{4}$$

We can now iterate (3) successively for $n = 2k$ and $n = 2k + 1$, using successive formulas for a_{2k+1} and a_{2k} and the fact $\log x = o(x^{\delta})$, to generate asymptotic expansions of a_n for odd n and even n. We will derive the first several terms. The first two steps yield

$$a_{2k} = 2k + O(k^{\varepsilon_2}) \quad \text{for} \quad \varepsilon_2 > \varepsilon_1 \tag{5}$$

and

$$a_{2k+1} = 1 + \frac{\log(2k+1)}{2k+1} + O(k^{\varepsilon_3 - 2}) \quad \text{for} \quad \varepsilon_3 > \varepsilon_2. \tag{6}$$

Having obtained two terms for a_{2k+1}, we can replace (4) by the more accurate formula $a_{2k+1} = 1 + O((\log k)/k)$. Resuming the iteration from that point, we replace (5) by $a_{2k} = 2k + O((\log(2k))^2)$ and (6) by

$$a_{2k+1} = 1 + \frac{\log(2k+1)}{2k+1} + O\left(\frac{(\log k)^3}{k^2}\right).$$

Finally, one more iteration yields

$$a_{2k} = 2k - \{\log(2k)\}^2 + O\left(k^{-1}(\log k)^4\right).$$

Editorial comment. Using the Maple symbolic computation package, Gaston H. Gonnet of the Symbolic Computation Group, University of Waterloo, Waterloo, Ontario, obtained the following more precise results. For n odd we have

$$a_n = 1 + \frac{\log n}{n} + \frac{2\log n + (\log n)^2 + 2(\log n)^3}{2n^2} + \frac{6\log n - 12(\log n)^2 + 25(\log n)^3 + 3(\log n)^4 + 9(\log n)^5}{6n^3} + O\left(\frac{1}{n^{4-\varepsilon}}\right),$$

while for n even we have

$$a_n = n - (\log n)^2 + \frac{2\log n - 4(\log n)^2 + (\log n)^3 - (\log n)^4}{2n} + \frac{15\log n - 30(\log n)^2 + 42(\log n)^3 - 25(\log n)^4 + 6(\log n)^5 - 4(\log n)^6}{6n^2} + O\left(\frac{1}{n^{3-\varepsilon}}\right).$$

All other solvers showed only that a_n approaches 1 for n tending to infinity through odd values and a_n/n approaches 1 for n tending to infinity through even values.

Solved also by L. Erdős (Hungary), J. Marengo, P. Tracy, and the editors. Partially solved by the proposer. Four incorrect solutions were received.

On Integer Triangles

E 3256 [1988, 259]. *Proposed by John Isbell, SUNY at Buffalo.*

(a) Let T be the set of triangles in the *plane* whose vertices have integral coordinates and whose sides have integral lengths. Certain isosceles triangles in T can be constructed by fitting together two congruent right triangles in T, e.g., the isosceles triangle with vertices $(-12, -9)$, $(12, 9)$, $(-12, 16)$ arises in this way. Are there any other isosceles triangles in T?

(b) Consider the set V of triangles in 3-*space* whose vertices have integral coordinates. Does V contain any equilateral triangles with integral side-length?

Composite solution by Jesús Ferrer, Oliva, Valencia, Spain, and Stephen M. Gagola, Jr., Kent State University, Ohio. (a) The answer is no. Let A, B, C be the vertices of a triangle in T with $AC = BC$. By translation, we may take $C = (0, 0)$. Let $A = (a_1, a_2)$, $B = (b_1, b_2)$, $x = AC = BC$, and $y = AB$. Then

$$x^2 = a_1^2 + a_2^2 = b_1^2 + b_2^2 \tag{1}$$

and

$$y^2 = (a_1 - b_1)^2 + (a_2 - b_2)^2 = 2\big(x^2 - a_1b_1 - a_2b_2\big), \tag{2}$$

Thus, y is even, so the segment of AB from A to its midpoint M has integer length. If y is even, then y^2 is $0 \bmod 4$, and congruence mod 4 implies that $a_1 - b_1$ and $a_2 - b_2$ are even. Hence $M = ((a_1 + b_1)/2, (a_2 + b_2)/2)$ has integer coordinates.

Finally, we compute CM. We may assume $a_2 \neq b_2$. By rearranging terms in (1), we obtain $(a_2 + b_2)/(a_1 + b_1) = -(a_1 - b_1)/(a_2 - b_2)$. Using this and then (2), we have

$$(CM)^2 = \left(\frac{a_1 + b_1}{2}\right)^2 + \left(\frac{a_2 + b_2}{2}\right)^2 = \left(\frac{a_1 + b_1}{2}\right)^2\left(1 + \left(\frac{a_2 + b_2}{a_1 + b_1}\right)^2\right)$$

$$= \left(\frac{a_1 + b_1}{2}\right)^2\left(1 + \left(\frac{a_1 - b_1}{a_2 - b_2}\right)^2\right) = \left(\frac{a_1 + b_1}{2}\right)^2 \frac{y^2}{(a_2 - b_2)^2}.$$

This is an integer and is the square of a rational number, so its square root must be an integer.

(b) The answer is no. Let A, B, C be the vertices of an equilateral triangle in V with side length x. By translation, we may take $C = (0, 0, 0)$. Let

$$A = (a_1, a_2, a_3), \qquad B = (b_1, b_2, b_3).$$

Without loss of generality we may assume that $a_1, a_2, a_3, b_1, b_2, b_3$ have no common factor greater than 1 and in particular that at least one of them is odd.

The equations used in (a) become

$$x^2 = a_1^2 + a_2^2 + a_3^2 = b_1^2 + b_2^2 + b_3^2$$

$$= (a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2 = 2\big(x^2 - a_1b_1 - a_2b_2 - a_3b_3\big). \tag{I}$$

Again we conclude that x is even, which implies

$$0 \equiv a_1^2 + a_2^2 + a_3^2 \equiv b_1^2 + b_2^2 + b_3^2 \pmod 4,$$

which is impossible when at least one of $a_1, a_2, a_3, b_1, b_2, b_3$ is odd.

Editorial comment. Several solvers noted that in higher dimensions the answers are different. In particular, the points $(0,0,0,0)$ $(2,0,0,0)$, and $(1,1,1,1)$ form the vertices of an equilateral triangle with integral side-length. If $n > 4$, we may append $n-4$ zero coordinates to give three points in Z^n forming the vertices of an equilateral triangle with integral side-length.

If F is a subfield of the reals, Jim Delany proved (a) there is an equilateral triangle with vertices in F^2 having side-length in F if and only if F contains $\sqrt{3}$, and (b) there is an equilateral triangle with vertices in F^3 having side-length in F if and only if 3 is a sum of two squares of elements of F.

I. J. Schoenberg's paper, "Regular simplices and quadratic forms," J. London Math. Soc. 12 (1937) 48–55 determines those values of n for which there exists an equilateral *n-simplex* whose vertices are in Z^n.

Also solved by J. Anglesio (France), J. Delany, K. G. Fischer, W. Janous (Austria), A. Pedersen (Denmark), A. J. Schwenk, and the proposer. Partially solved by J.-M. Monier (France).

A Special Collection of Permutations

E 3260 [1988, 350]. *Proposed by Peter Andrews and Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Ontario, Canada.*

In how many ways can n white squares and n black squares be chosen from a $2n$ by $2n$ chessboard in such a way that no two of the chosen squares lie in the same row or the same column?

Solution and generalization by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. If n is odd there are no such permutations, but if n is even the answer is $[n!/(n/2)!]^4$.

Translate a permissible choice into a permutation π such that choosing position (i, j) corresponds to $\pi(i) = j$. The squares of one color are those with $i \equiv j \bmod 2$; the other color have $i \neq j \bmod 2$. Thus the desired set corresponds to the permutations of $\{1,\ldots,2n\}$ such that $i \not\equiv \pi(i) \bmod 2$ for exactly n values of i. More generally, let $f(m, k)$ denote the number of permutations of $[m] = \{1,\ldots, m\}$ such that $i \not\equiv \pi(i) \bmod 2$ for exactly k values of i; we compute $f(m, k)$ for all $0 \leqslant k \leqslant m$.

Note first that in any permutation π of $[m]$, the number of odd values that have their parity changed must equal the number of even values that have their parity changed, because the total number of odds must be the same before and after the permutation. Hence $f(m, k) = 0$ if k is odd.

If k is even, the desired permutations can be constructed by choosing the numbers whose parities will be reversed and then choosing a permutation for the $\lceil m/2 \rceil$ numbers that will be mapped to odd numbers and a permutation for the $\lfloor m/2 \rfloor$ numbers that will be mapped to even numbers. The number of ways to do this is

$$f(m,k) = \binom{\lceil m/2 \rceil}{k/2}\binom{\lfloor m/2 \rfloor}{k/2}\lceil m/2 \rceil!\lfloor m/2 \rfloor!$$

When $m = 2n$ and $k = n$, this reduces to the formula claimed.

Editorial comment. The proposers remarked that, since the total number of permutations is $m!$, the explicit determination of $f(m, k)$ yields the interesting but well-known combinatorial identity $\sum_{k=0}^{m} f(m, k) = m!$.

Solved also by the proposers and 24 others.

Simplices with Integer Edges and Volume One

E 3279 [1988, 655]. *Proposed by Stanley Rabinowitz, Westford, MA.*

For some n ($n > 1$), find a simplex in E^n with integer edges and volume 1. (Cf. Problem 1261 in *Mathematics Magazine* [1987, p. 40; 1988, p. 118].)

Remark. The condition $n > 1$ was inadvertently omitted from the published statement of the problem.

Solution by the proposer. In E^4, the simplex with vertices

$$(0,0,0,0),\quad (7,0,0,0),\quad \left(33/14, 15\sqrt{3}/14, 0, 0\right),$$
$$\left(81/14, -199\sqrt{3}/70, \sqrt{582}/5, 0\right),$$
$$\left(53/7, 652\sqrt{3}/315, 1288\sqrt{582}/4365, 8\sqrt{194}/291\right)$$

has the desired property.

Editorial comment. As Walther Janous (Austria) remarked, Heron's formula shows immediately that a triangle with integer edges cannot have area 1. Actually it is known that if a triangle has integer edges and integer area, then the area must be a multiple of 6. (See the above-mentioned problem 1261 in the *Mathematics Magazine*.) In any event, a simplex of the type mentioned in the statement of the problem does not exist for $n = 2$.

In a paper, "Tetrahedrons with integer edges and integer volume," to appear in *Mathematics Magazine*, Kevin L. Dove and John S. Sumner show that there is no tetrahedron with integer edges and volume 1. Thus $n = 4$ is the smallest value of n for which the phenomenon of the problem occurs.

It is conceivable that a numerically simpler example than that given above can be found for some larger value of n. However, the editors know of no such configuration.

No other solutions were received.

ADVANCED PROBLEMS

6619. *Proposed by Arthur L. Holshouser, Charlotte, NC, and Benjamin G. Klein, Davidson College, NC.*

Suppose that $N = 4^h(8k + 7)$ for non-negative integers h and k.

(a) Show that there exist infinitely many distinct 5-tuples (a, b, c, d, e) of positive integers such that

$$N = (a^2 + b^2 + c^2 + d^2 + e^2)/(abcde + 1).$$

(b) Show that there is no 4-tuple (a, b, c, d) of positive integers such that

$$N = (a^2 + b^2 + c^2 + d^2)/(abcd + 1).$$

6620. *Proposed by W. Allegretto and A. Meir, University of Alberta, Edmonton, Canada.*

Suppose f is locally integrable on R, but

$$\int_R |f(x)|\,dx = \infty.$$

Show that for every $p > 0$ there exists $\phi \in L^p(R)$ such that it is impossible to have

$$f(x)\phi(x) = g(x)h(x) \qquad (x \in R),$$

where $g \in L^1(R)$ and $h \in L^p(R)$.

6621. *Proposed by William C. Waterhouse, Pennsylvania State University, University Park.*

Which numbers can occur as a ratio of roots of an irreducible quadratic equation? More precisely, suppose K is an algebraically closed field. For each subfield E of K and each separable quadratic extension F of E with conjugation $\bar{}$, certain elements of K occur as ratios $\bar{c}/c$ with c in F. Which elements of K (besides 1) do arise in this way when we allow all possible choices of E and F?

SOLUTIONS OF ADVANCED PROBLEMS

Ratios of the Terms of a Sequence

6559 [1987, 885]. *Proposed by Benjamin G. Klein, Davidson College, and John Layman, Virginia Polytechnic Institute and State University.*

Let $c_0, c_1, c_2, \ldots$ be a sequence of real numbers determined by the values of c_0, c_1, c_2 together with the recursion

$$c_{n+1} = (c_n c_{n-1} + 1)/c_{n-2},$$

for $n \geq 2$. Further, suppose that $c_0 = c_2 = a$, where $a > 0$.

(i) For what values of c_1 does the sequence of ratios c_{2n+2}/c_{2n} converge? Determine the value of the limit for each such value of c_1.

(ii) Show that for each value of a there is exactly one value of c_1, such that the sequence of ratios c_{n+1}/c_n converges. Find this value of c_1 and then determine the associated value of the limit of the sequence of ratios.

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. We shall show that c_{2n+2}/c_{2n} has a limit only in the following three cases:

$$\text{I.}\quad c_1 = -1/(2a). \qquad \text{II.}\quad -1/(2a) < c_1 < 0. \qquad \text{III.}\quad c_1 > 0.$$

In case I the limit is -1, in case II the limit is $(B - \sqrt{B^2 - 4})/2$, where $B = 2 + 2/(c_1 a)$, and in case III the limit is $(B + \sqrt{B^2 - 4})/2$.

The ratio c_{n+1}/c_n has a limit only if $c_1 = (-1 + \sqrt{1 + 16a^4})/(4a)$ and then the value of the limit is

$$\left(2ac_1 + 1 + \sqrt{2ac_1 + 1}\right)/(2a^2).$$

We now proceed to prove these assertions. Note that the recursion breaks down if any of the numbers c_i vanishes. So let us assume that at every stage of the recursion the number c_j in the denominator does not vanish and check afterwards whether all c_j do not vanish. From the initial elements $(c_0, c_1, c_2) = (a, c, a)$ and the recursion it follows that

$$c_3 = (ca + 1)/a \quad \text{and} \quad c_4 = (ca + 2)/c.$$

Claim. If $c_n = Bc_{n-2} - c_{n-4}$ for some B, then $c_{n+1} = Bc_{n-1} - c_{n-3}$.

Proof.

$$\begin{aligned} c_{n+1}c_{n-2} &= c_nc_{n-1} + 1 = (Bc_{n-2} - c_{n-4})c_{n-1} + 1 \\ &= Bc_{n-2}c_{n-1} - (c_{n-3}c_{n-2} + 1) + 1 \\ &= (Bc_{n-1} - c_{n-3})c_{n-2}. \end{aligned}$$

Clearly in our situation $B = (c_4 + c_0)/c_2 = 2 + 2/(ca)$. We now distinguish three cases.

(1) $B = 2$ is impossible.

(2) $B = -2$ yields $ca = -1/2$, so $(c_0, c_2, c_4) = (a, a, -3a)$, hence $c_{2n} = a(1 - 2n)(-1)^n$, and $(c_1, c_3) = (c, -c)$, hence $c_{2n+1} = c(-1)^n$.
Clearly all $c_j \neq 0$, the ratio c_{n+2}/c_n tends to -1 (in both cases), and c_{n+1}/c_n has no limit.

(3) If $|B| \neq 2$, then

$$c_{2n} = \alpha\lambda^n + \beta\mu^n; \qquad c_{2n+1} = \gamma\lambda^n + \delta\mu^n,$$

where λ and μ are the roots of the equation

$$x^2 - Bx + 1 = 0. \tag{*}$$

Using the starting values we obtain

$$\begin{cases} c_{2n} = a\left(\dfrac{\lambda^n}{\lambda + 1} + \dfrac{\mu^n}{\mu + 1}\right), \\ \text{and} \\ c_{2n+1} = \dfrac{1}{2}c(\lambda^n + \mu^n). \end{cases} \tag{**}$$

We now subdivide case (3) into three sub-cases.

a) If $|B| < 2$ then λ and μ are complex conjugates, so c_{n+2}/c_n has no limit.

b) If $B < -2$ then (*) has two distinct negative roots, so c_{n+2}/c_n tends to the smaller root, i.e.,

$$c_{n+2}/c_n \to \frac{1}{2}(B - \sqrt{B^2 - 4}) < 0,$$

and c_{n+1}/c_n has no limit.

c) If $B > 2$ then (*) has two positive roots, so

$$c_{n+2}/c_n \to \frac{1}{2}(B + \sqrt{B^2 - 4}) > 0.$$

We now verify that for $|B| > 2$ none of the numbers c_n can be zero. Since $\lambda = \mu^{-1}$, and without loss of generality $|\lambda| > |\mu|$, we have

$$|c_{2n}| = \frac{|a|}{|\lambda + 1|}|\lambda^n + \lambda\mu^n| \geqslant \frac{|a|\,|\lambda|}{|\lambda + 1|}(|\lambda|^{n-1} + |\mu|^n) > 0$$

and, trivially,

$$|c_{2n+1}| \geqslant \left|\frac{c}{2}\right|(|\lambda|^n - |\mu|^n) > 0.$$

For (ii) we may, because of the foregoing arguments, restrict ourselves to the case $B > 2$, and assume $\lambda > 1 > \mu$. From (**) it follows that

$$\frac{c_{2n+2}}{c_{2n+1}} = \frac{2a}{c}\left(\frac{\lambda^{n+1}}{\lambda + 1} + \frac{\mu^{n+1}}{\mu + 1}\right)\Big/(\lambda^n + \mu^n) \to \frac{2a}{c}\frac{\lambda}{\lambda + 1}.$$

In the same way

$$\frac{c_{2n+1}}{c_{2n}} \to \frac{c}{2a}(\lambda + 1) \quad \text{as } n \to \infty.$$

In order for c_{n+1}/c_n to have a limit as $n \to \infty$ we, therefore, must require that

$$\frac{2a}{c}\frac{\lambda}{\lambda + 1} = \frac{c}{2a}(\lambda + 1);$$

simultaneously with (*), that is,

$$\lambda^2 - \left(2 + \frac{2}{ac}\right)\lambda + 1 = 0.$$

These equations have a root λ in common (and are even identical) if

$$2 - 4a^2/c^2 = -2 - 2/(ac).$$

This means, since c must be positive, that

$$c = \frac{-1 + \sqrt{1 + 16a^4}}{4a} \qquad (***)$$

and

$$\frac{c_{n+1}}{c_n} \to \frac{c}{2a}(\lambda + 1) = \frac{1}{2a^2}(2ac + 1 + \sqrt{2ac + 1}),$$

where c is determined by (***).

Also solved by C. K. Bailey and R. B. Richter, Robert J. Currier, Lorraine L. Foster, Ha-seo Ki (Korea), D. Hensley, Kim Jun Kyo (Korea), R. J. Tomkins, and the proposers.

Limit Involving a Power of the Sine

6564 [1988, 60]. *Proposed by Gerald Weinstein, City College, City University of New York.*

Define $S(\theta) = \lim_{n \to +\infty}|\sin n|^{n^\theta}$.

(i) Show that $S(2)$ does not exist.

(ii) If C is a constant such that $|p/q - \pi| > q^{-C}$ for all pairs p, q of sufficiently large positive integers, prove that $S(\theta)$ exists and is zero for all $\theta > 2C - 2$.

Mahler [*Indagationes Math*, 15 (1953) 30–42] showed that we can take $C = 41.09$. For the best known value of C see D. V. Chudnovsky and G. V. Chudnovsky, *Springer Lecture Notes 1052*, (1984), 37–84.

Editorial remark. The absolute value symbols around $\sin\theta$ were unfortunately omitted from the definition of $S(\theta)$ in the published statement of the problem.

Solution by Ha-seo Ki, student, Yonsei University, Seoul, Korea and the proposer (combined). It is well known that for any positive irrational number β there are infinitely many pairs of relatively prime positive integers n and k such that

$$\left|\beta - \frac{n}{k}\right| < \frac{c_0}{k^2}, \qquad 0 < c_0 < 1 \tag{*}$$

where c_0 is an absolute constant. Hence

$$|2\pi k - n| < \frac{1}{k}$$

infinitely often, and we see that for $\theta \geqslant 0$ the limit (if it exists) must be 0. We also note that for sufficiently small angles ϕ we have

$$1 - \phi^2 \leqslant \cos\phi \leqslant 1 - \frac{\phi^2}{3}.$$

Fix $\varepsilon > 0$. We now ask if there are infinitely many pairs of integers N and K (call them "epsilon" pairs) such that

$$\left|2\pi K + \frac{\pi}{2} - N\right| < \left(1 - \varepsilon^{N^{-\theta}}\right)^{1/2}.$$

If so, for large N we would have

$$\begin{aligned}\sin N &= \cos\left(\lambda\left(1 - \varepsilon^{N^{-\theta}}\right)^{1/2}\right) \qquad (0 \leqslant \lambda < 1)\\ &\geqslant 1 - \left(1 - \varepsilon^{N^{-\theta}}\right) = \varepsilon^{N^{-\theta}}\end{aligned}$$

and the limit defining $S(\theta)$ could not exist.

Now let β be a positive irrational number, σ any real number, and n, k a typical pair approximating β as in the first paragraph. Let t be the integer nearest to $k\sigma$, so

$$\sigma = \frac{t}{k} + \frac{c_1}{2k}, \qquad |c_1| \leqslant 1.$$

For integers K and N we have

$$\begin{aligned}|\beta K - N - \sigma| &= \left|\left(\frac{n}{k} + \frac{c_2}{k^2}\right)K - N - \frac{t}{k} - \frac{c_1}{2k}\right| \qquad (|c_2| \leqslant 1)\\ &= \left|\frac{nK - kN - t}{k} + \frac{c_2 K}{k^2} - \frac{c_1}{2k}\right|.\end{aligned}$$

Since n and k are relatively prime we may choose K and N so that the first fraction

above is 0, and in fact with $k \leqslant K \leqslant 2k$. Thus, if k is sufficiently large,

$$|\beta K - N - \sigma| \leqslant \frac{2k}{k^2} + \frac{1}{2k} = \frac{5}{2k} \leqslant \frac{5}{K} < \frac{c_3(\beta, \sigma)}{N}. \qquad (**)$$

Since there are infinitely many pairs k, n satisfying $(*)$, it follows that $(**)$ holds for infinitely many pairs K, N. (A more careful estimate due to Khintchine shows that the 5 may be replaced by $5^{-1/2} + \eta$ for any $\eta > 0$).

For $\beta = 2\pi$ and $\sigma = -\pi/2$ it follows that there are infinitely many "epsilon" pairs if

$$\frac{c_3}{N} < \left(1 - \exp\left(N^{-\theta} \log \varepsilon\right)\right)^{1/2} \sim N^{-\theta/2}\left(\log \frac{1}{\varepsilon}\right)^{1/2}$$

for all large N. Clearly for $\theta = 2$ and ε a fixed but sufficiently small positive number this is valid. Thus (i) is established.

For (ii) the simple idea is to show that n cannot get too close to $\pi/2$ or $3\pi/2$ modulo 2π infinitely often; we give the details only for $\pi/2$. For n large,

$$\begin{aligned} a(n) = \left|2\pi k + \frac{\pi}{2} - n\right| &= \left(2k + \frac{1}{2}\right)\left|\pi - \frac{2n}{4k+1}\right| \\ &> \left(2k + \frac{1}{2}\right)(4k+1)^{-C} =: b, \end{aligned}$$

where b is very small. We may confine our attention to the case $a(n) < 1/10$. It is then obvious that

$$|\sin n| \leqslant \cos b$$

and thus, that there are positive constants c_4, c_5, and c_6 such that

$$\begin{aligned} |\sin n| &\leqslant \cos\left(c_4 k^{1-C}\right) \leqslant \cos\left(c_4 n^{1-C}\right) \\ &\leqslant 1 - \frac{c_6}{n^{2C-2}}. \end{aligned}$$

Hence,

$$\begin{aligned} |\sin n|^{n^\theta} &\leqslant \left[\left(1 - \frac{c_6}{n^{2C-2}}\right)^{n^{2C-2}}\right]^{n^{\theta-(2C-2)}} \\ &\leqslant \exp\left(-c_6 n^{\theta-(2C-2)}\right), \end{aligned}$$

which, for $\theta > 2C - 2$, tends to 0 as n tends to infinity. In this way (ii) is established.

No other solutions were received.

Sequences Containing Arbitrarily Long Arithmetic Progressions

6565. *Proposed by L. A. Rubel, University of Illinois at Urbana-Champaign.*

Suppose $A = \{a_1, a_2, a_3, \dots\}$ is a sequence of distinct positive integers and $B = \{b_1, b_2, b_3, \dots\}$ is another. Suppose A contains arbitrarily long arithmetic progressions.

(a) If $|b_n - a_n| \leqslant K$ for all n, where K is a fixed positive number, prove that B must contain arbitrarily long arithmetic progressions.

(b) If $\Sigma_{n=1}^{\infty} 1/\{1 + |b_n - a_n|\}$ diverges, does the same conclusion hold?

Solution by Joseph E. McCanna, University of New Mexico, Albuquerque, NM and Douglas Hensley, Texas A & M University, College Station, TX (combined). We shall prove (a) and then show that no such conclusion is valid for (b). Fix $M > 0$. Let $g_1 < g_2 < \cdots < g_L$ be an arithmetic progression of length L in A. "Color" each $g_l = a_{n_l}$ with the number $b_{n_l} - a_{n_l}$. By hypothesis we are only using a finite number of colors. By van der Waerden's theorem there is a monochromatic arithmetic progression of length $N \geqslant M$ among the g_l if L is large enough. But clearly this means there is also an arithmetic progression of length N in the sequence B. Next, let $\{a_n\}$ be the sequence

$$10, 11, 1000, 1001, 1002, 10^6, \ldots, 10^6 + 3, 10^{10}, \ldots, 10^{10} + 4, \ldots$$

and $\{b_n\}$ the sequence

$$10, 10^2, 10^3, 10^4, 10^5, 10^6, \ldots$$

Since $a_n = b_n$ for $n = 1, 3, 6, 10, 15, \ldots$ it is clear that the series diverges. Clearly A has arbitrarily long arithmetic progressions, while B has none.

Also solved by Alan H. Forrest (student), Hillel Gauchman, Robert B. Israel (Canada), Bruce Reznick, Lior Tzafriri, and Daniel Ullman.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Basic Hypergeometric Series and Applications. By Nathan J. Fine, American Mathematical Society, 1988. xiii + 124 pp. ISBN 0-8218-1524-5. $39.00

RANJAN ROY
Department of Mathematics and Computer Science, Beloit College, Beloit, WI 53511

A basic hypergeometric series is written as

$$ {}_m\Phi_n\left[\begin{matrix} a_1,\ldots,a_m;\ q;\ z \\ b_1,\ldots,b_n. \end{matrix}\right] = \sum_{k\geqslant 0} \frac{(a_1)_k(a_2)_k\cdots(a_m)_k z^k}{(b_1)_k(b_2)_k\cdots(b_n)_k(q)_k}, \tag{1} $$

where $(a)_0 = 1$ and $(a)_k = (a;\ q)_k = (1-a)(1-aq)\cdots(1-aq^{k-1})$ for $k \geqslant 1$ and $|z| < 1$, $|q| < 1$, $b_i \neq q^{-n}$ for any nonnegative integer n. We get a bilateral series if we allow k to assume negative values as well:

$$ {}_m\Psi_n\left[\begin{matrix} a_1,\ldots,a_m;\ q;\ z \\ b_1,\ldots,b_n. \end{matrix}\right] = \sum_{k=-\infty}^{\infty} \frac{(a_1)_k(a_2)_k\ldots(a_m)_k z^k}{(b_1)_k(b_2)_k\ldots(b_n)_k}, \tag{2} $$

where $(a)_{-k} = (1-a/q^k)^{-1}(1-a/q^{k-1})^{-1}\cdots(1-a/q)^{-1}$. We can obtain Jacobi's theta series from (2), for example,

$$ \sum_{k=-\infty}^{\infty} q^{k^2}z^k \tag{3} $$

is found by setting $n = 0$, $m = 1$, $a_1 = (1/t)$ and replacing q by q^2 and z by $-tqz$ and then setting $t = 0$. Basic or q-series have important applications to the theory of partitions, the representation of numbers as sums of squares, the classification and proofs of binomial identities, statistical mechanics, and other areas of mathematics.

The earliest examples of these series can be found in the work of L. Euler. He first conjectured in 1741 and later proved in 1750 the beautiful formula:

$$ \prod_{n=0}^{\infty}(1-q^n) = \sum_{n=-\infty}^{\infty}(-1)^n q^{(1/2)n(3n+1)}, \tag{4} $$

where the numbers $(1/2)n(3n+1)$ are called pentagonal numbers. The series in (4) is a particular case of (3) obtained by replacing q^2 by q^3 and z by $-q^{1/2}$. Euler was led to consider the product in (4) after his discovery that

$$ \prod_{n=0}^{\infty}(1-q^n)^{-1} = \sum_{n=0}^{\infty} p(n)q^n, \tag{5} $$

where $p(n)$ denotes the number of unrestricted partitions of n, that is, the number of ways in which n can be written as a sum of numbers less than or equal to n. A. M. Legendre discovered the following combinatorial interpretation of (4): Let $p_e(n)$ (resp. $p_o(n)$) denote the number of partitions of n into an even (resp. odd)

number of distinct parts. Then (4) implies that

$$p_e(n) - p_o(n) = \begin{cases} (-1)^m & \text{if } n = \frac{1}{2}m(3m \pm 1) \\ 0 & \text{otherwise.} \end{cases}$$

Euler was able to use (4) and (5) to obtain a formula which provides an efficient algorithm for computing $p(n)$, namely,

$$p(n) - p(n-1) - p(n-2) + p(n-5) + p(n-7) + \cdots + (-1)^m p\left(n - \frac{1}{2}m(3m-1)\right) + (-1)^m p\left(n - \frac{1}{2}m(3m+1)\right) + \cdots = 0, \tag{6}$$

where it is understood that $p(k) = 0$ for k negative.

C. F. Gauss encountered the very important basic series $1 + 2\Sigma_{n=1}^{\infty} q^{n^2}$ and $1 + 2\Sigma_{n=1}^{\infty}(-1)^n q^{n^2}$, which are particular cases of (3), in the earliest stage of his mathematical career. There is an oral tradition that he had observed the connection of these series with the arithmetic-geometric mean in 1791 when he was only fourteen. It was somewhat later in the period 1797–1800 that he found the relation of these topics with the theory of elliptic and modular functions and the division of an arc of a lemniscate into equal parts. These were very great discoveries but he published almost nothing of them. He also summed the above mentioned series as infinite products:

$$1 + 2\sum_{n=1}^{\infty} q^{n^2} = \prod_{n=1}^{\infty}(1 - q^{2n})(1 + q^{2n-1})^2, \tag{7}$$

$$1 + 2\sum_{n=1}^{\infty} (-1)^n q^{n^2} = \prod_{n=1}^{\infty}(1 - q^{2n})(1 - q^{2n-1})^2. \tag{8}$$

After he found these results Gauss set himself the problem of expressing $1 + q + q^3 + q^6 + q^{10} + \cdots$, a series whose powers are triangular numbers rather than squares, as an infinite product. He derived the answer that

$$1 + \sum_{n=1}^{\infty} q^{n(n+1)/2} = \prod_{n=1}^{\infty} \frac{1 - q^{2n}}{1 - q^{2n-1}} \tag{9}$$

from his theory of elliptic modular functions. However, in 1811 he published another derivation independent of elliptic functions. The ideas of this paper are important in the study of q-series. It may be of interest to note that this paper also contains his evaluation of the sum $1 + \Sigma_{k=1}^{n} e^{2k^2\pi i/n}$, which was of crucial importance in his fourth proof of the quadratic reciprocity law. The problem of its evaluation had worried Gauss for almost four years (1801–1805). A history of Gauss' work on elliptic and modular functions is contained in an interesting paper by D. Cox [**9**].

In a 1750 letter to Goldbach, Euler suggested that the most natural way of proving the four-square theorem would be to expand $(1 + q + q^4 + q^9 + \cdots)^4$ as a power series $\Sigma_{n=0}^{\infty} c_n q^n$ and show that $c_n \neq 0$. In fact c_n would also give the number of representations of a positive integer as a sum of four squares. Euler had

no clue as to how this could be done but in 1828 C. Jacobi succeeded in proving the following identities:

$$(1 + 2q + 2q^4 + \cdots)^2 = 1 + 4\sum_{n=0}^{\infty} \frac{(-1)^n q^{2n+1}}{1 - q^{2n+1}} \tag{10}$$

$$(1 + 2q + 2q^4 + \cdots)^4 = 1 + 8 \sum_{\substack{n \not\equiv 0 \bmod 4 \\ n \geqslant 1}} \frac{nq^n}{1 - q^n}. \tag{11}$$

It is an easy consequence of the first identity (10) that the number of ways of expressing a positive integer n as a sum of two squares, (counting order and sign) is four times the difference of the number of divisors of n congruent to 1 modulo 4 and the number of those congruent to 3 modulo 4. The second identity (11) gives the number of representations of n as a sum of four squares to be

$$8 \sum_{\substack{d/n \\ d \not\equiv 0 \bmod 4}} d.$$

Jacobi derived these identities from his theory of elliptic functions. As another consequence of his theory, Jacobi found the famous triple product identity

$$\sum_{n=-\infty}^{\infty} q^{n^2} z^n = \prod_{n=0}^{\infty} (1 - q^{2n+2})(1 + zq^{2n+1})(1 + z^{-1}q^{2n+1}), \tag{12}$$

which contains (4), (7), (8), and (9) as particular cases. Gauss too had discovered (12) independently. There are several elementary proofs of (12) which do not use elliptic functions. One such proof is in the book under review. M. Hirschhorn [**13**, **14**] has recently shown that Jacobi's identities (9) and (10) on sums of squares can be directly obtained from (12) by the elementary processes of differentiation and algebraic manipulation of series. An excellent treatment of these matters is contained in [**7**]. G. Andrews [**2**] has found that there is an identity similar to (10) and (11) for the cube of the series in (9) from which one can obtain Gauss' theorem that every number is the sum of three triangular numbers.

The first person to study basic hypergeometric series in a general way was E. Heine. Euler, Gauss and Jacobi had studied only those series which arise in modular function theory. In 1847, Heine published the following transformation formula:

For $|q| < |$, $|z| < |$, and $|b| < 1$

$$\sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(q)_n (c)_n} z^n = \frac{(b)_\infty (az)_\infty}{(c)_\infty (z)_\infty} \sum_{n=0}^{\infty} \frac{(c/b)_n (z)_n}{(q_n)(az)_n} b^n, \tag{13}$$

where $(x)_\infty = \prod_{n=0}^{\infty}(1 - xq^n)$. A particular case of this is the "q-analog of the binomial theorem":

$$\sum_{n=0}^{\infty} \frac{(a)_n}{(q)_n} z^n = \frac{(az)_\infty}{(z)_\infty}, \tag{14}$$

a fundamental result in the theory of basic series, which was known to Gauss and Cauchy though it appears that H. Rothe was the first to publish it in 1811. The binomial theorem follows from it if we set $a = q^\alpha$ and let $q \to 1$. Particular cases of

(14) were known to Euler:

$$\sum_{n=0}^{\infty} \frac{z^n}{(q)_n} = \frac{1}{(z)_\infty}, \tag{15}$$

$$\sum_{n=0}^{\infty} \frac{z^n q^{n(n-1)/2}}{(q)_n} = (-z)_\infty. \tag{16}$$

From the general theorem of Heine one can easily obtain as a particular case the following identity of Cauchy:

$$\sum_{n=0}^{\infty} \frac{q^{n^2-n} z^n}{(q)_n (z)_n} = \frac{1}{(z)_\infty}, \tag{17}$$

and if we set $z = q$ we get a formula of Euler:

$$1 + \sum_{n=1}^{\infty} \frac{q^{n^2}}{(1-q)^2 \cdots (1-q^n)^2} = \prod_{m=1}^{\infty} (1-q^m)^{-1}. \tag{18}$$

We also note that if we replace q by q^2 and set $z = q$ in (16) we obtain

$$1 + \sum_{n=1}^{\infty} \frac{q^{n^2}}{(1-q^2) \cdots (1-q^{2n})} = \prod_{m=0}^{\infty} (1+q^{2m-1}). \tag{19}$$

These identities have interesting combinatorial interpretations. For this as well as for recent developments in this area the reader may enjoy seeing Andrews [**3**].

The identities (18) and (19) are not very difficult to prove but the following which are superficially similar are extremely hard:

$$1 + \sum_{n=1}^{\infty} \frac{q^{n^2}}{(1-q) \cdots (1-q^n)} = \prod_{m=0}^{\infty} (1-q^{5m+1})^{-1} (1-q^{5m+4})^{-1}, \tag{20}$$

$$1 + \sum_{n=1}^{\infty} \frac{q^{n(n+1)}}{(1-q) \cdots (1-q^n)} = \prod_{m=0}^{\infty} (1-q^{5m+2})^{-1} (1-q^{5m+3})^{-1}. \tag{21}$$

These are due to L. J. Rogers and S. Ramanujan. They have an interesting history which has been told many times but may still be worth repeating. Ramanujan discovered these identities around 1910 but had no proof. He communicated them to G. H. Hardy in February, 1913 but neither Hardy nor others consulted by Hardy could find a proof. P. MacMahon expanded both sides as a series and verified the equality up to 89 terms. This convinced him of the truth of the results which he included without proof in his book on combinatorics [**15**]; a book which is even now of more than historic interest. MacMahon found the following interpretation of the first identity: *The number of partitions of n with minimal difference* 2 *is equal to the number of partitions into parts of the form* $5m + 1$ *or* $5m + 4$. The second identity has a similar interpretation.

In 1916, Ramanujan accidentally came across Rogers' 1894 paper in the *Proceedings of the London Mathematical Society* which contained the statements and proofs of the identities. As a result Rogers and his work theretofore completely neglected became known. However, it was only recently that the greater significance of Rogers' work was understood. The reader may wish to consult [**2**] for details of this

interesting story. The two identities were also rediscovered by I. Schur in 1917 and then in 1979 by R. Baxter who utilized them in his solution of the Hard Hexagon Model [**6**]. Some interesting mathematics has arisen from reproving and generalizing these identities. A purely combinatorial proof involving a new convolution principle was found by A. Garsia and S. Milne [**10**] though the simplest proof is perhaps the one due to D. Bressoud [**8**]. No proof depending on modular functions is known.

Ramanujan used the two identities as the basis for his theory of continued fractions of the form

$$\frac{1}{1+}\,\frac{x}{1+}\,\frac{x^2}{1+}\cdots.$$

It is not difficult to show that this continued fraction is the ratio of the two series in the identities. He used this to obtain some remarkable formulae, such as

$$\frac{1}{1+}\,\frac{e^{-2\pi}}{1+}\,\frac{e^{-4\pi}}{1+}\cdots = \left\{\sqrt{\left(\frac{5+\sqrt{5}}{2}\right)} - \frac{\sqrt{5+1}}{2}\right\} e^{2\pi/5}. \tag{22}$$

Hardy's comment on these results was, "I had never seen anything in the least like them before. A single look is enough to show that they could only be written by a mathematician of the highest class." A coverage of this aspect of Ramanujan's work is contained in [**1**].

Ramanujan extended the theory of basic series in many directions. Apparently he was the first to sum the bilateral series

$${}_1\psi_1(a;\, b;\, z) = \sum_{n=-\infty}^{\infty} \frac{(a)_n z^n}{(b)_n} = \frac{(b/a)_\infty (az)_\infty (a/az)_\infty (q)_\infty}{(q/a)_\infty (b/az)_\infty (b)_\infty (z)_\infty}. \tag{23}$$

This result was somehow missed by Gauss and Jacobi. It contains the triple product identity and the q-binomial theorem as special cases. It arises sufficiently often in mathematics to be considered an important result. Ramanujan's identity was extracted from his notebooks by Hardy (see [**12**]), who refers to it as "a remarkable formula with many parameters." Several proofs of this formula have appeared and one is contained in the book under review. R. Askey's [**5**] observation that (23) and some other formulae are in fact generalizations or analogs of Euler's beta integral

$$\int_0^1 \frac{x^{\alpha-1}}{(1+x)^{\alpha+\beta}}\,dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

throws new and fruitful light on them. One gets some conception of the origin of such formulae and what can be done with them.

In the last year of his life, Ramanujan introduced a new kind of q-series into mathematics which he named "mock-theta functions." He wrote about them in a letter to Hardy two months before his death. G. N. Watson developed the theory of these functions on the basis of that letter and apparently it even motivated the young A. Selberg to do some early work in this area. However, it was not until very recently that mock-theta functions began to play a greater role in mathematics,

inspired by the dramatic discovery of Andrews of a lost notebook of Ramanujan. Mock-theta functions may yet be recognized as Ramanujan's greatest contribution to mathematics.

If in (1) we set $m = n + 1$ and substitute q^{a_i} for a_i and q^{b_i} for b_i and let $q \to 1$ we get the ordinary hypergeometric series

$$\sum_{k \geq 0} \frac{(a_1)_k (a_2)_k \cdots (a_m)_k z^k}{(b_1)_k (b_2)_k \cdots (b_n)_k (1)_k}, \tag{24}$$

where $(a)_k = a(a + 1) \cdots (a + k - 1)$. These series have played an extraordinarily important role in mathematics and physics and were studied extensively by Euler, Gauss, Kummer, Riemann, Ramanujan, and numerous other mathematicians. Recently they played a significant part in L. DeBrange's solution of the Bieberbach conjecture. It is also now generally becoming known that they can be used to bring some order into the problem of classifying binomial identities. For this, the reader may consult [**4**], [**11**], and [**16**].

N. J. Fine's book provides a fascinating introdution to the subject of basic or q-hypergeometric series. We see that it is possible to obtain beautiful and deep mathematical results without the development of a heavy and elaborate machinery. In fact, this book can be studied very profitably by a talented undergraduate. From the introduction to the book we learn that in 1948 Fine himself embarked upon an investigation of basic series with only a modicum of casual information about the subject. He was led to several elegant new results on partitions and a rediscovery of many important old results, among them the evaluation of Ramanujan's ${}_1\psi_1$ sum. Over the years the author has added to his results and polished his presentation and this book is a consequence of this labor of love. Topics covered in the book include partitions, representation of numbers as sums of squares, mock theta functions and modular equations, all of which are subjects of great research interest today. G. Andrews has added notes at the end of each chapter giving references to some of the most recent developments. Almost every mathematician should find something to attract or charm her/him in this book. It is highly recommended.

REFERENCES

1. C. Adiga, B. C. Berndt, S. Bhargava, and G. N. Watson, Ramanujan's second notebook: Theta functions and q-series, chapter 16, Mem. Amer. Math. Soc., No. 315, 1985.
2. G. Andrews, q-series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra, Amer. Math. Soc., 1986.
3. _____, The Theory of Partitions, Encyclopedia of Mathematics and Its Applications, Vol. 2, Addison-Wesley, Reading, 1976.
4. _____, Applications of basic hypergeometric functions, *SIAM Review*, 16 (1974) 441–484.
5. R. Askey, Ramanujan's extensions of the gamma and beta functions, *Amer. Math. Monthly*, 87 (1980) 346–359.
6. R. Baxter, Rogers-Ramanujan identities in the hard hexagon model, *J. Statist. Phys.*, 26 (1981) 427–452.
7. J. Borwein and P. Borwein, Pi and the AGM, J. Wiley, New York, 1987.
8. D. Bressoud, An easy proof of the Rogers-Ramanujan identities, *J. Number Theory*, 16 (1983) 235–241.
9. D. Cox, The arithmetic-geometric mean of Gauss, *L'Enseignement Math.*, 30 (1984) 275–330.
10. A. Garsia and S. Milne, A Rogers-Ramanujan bijection, *J. Combin. Theory Ser A*, 31 (1981) 289–339.

11. R. Graham, D. Knuth, and O. Patashnik, Concrete Mathematics, Addison-Wesley, Reading, 1989.
12. G. H. Hardy, Ramanujan, Chelsea, New York, 1978.
13. M. Hirschhorn, A simple proof of Jacobi's two square theorem, *Amer. Math. Monthly* 92 (1985) 579–580.
14. _____, A simple proof of Jacobi's four square theorem, *Proc. Amer. Math. Soc.*, 101 (1987) 436–438.
15. P. MacMahon, Combinatory Analysis, Cambridge University Press, Cambridge, 1916.
16. R. Roy, Binomial identities and hypergeometric series, *Amer. Math. Monthly*, 94 (1987) 36–46.

Innumeracy: Mathematical Illiteracy and Its Consequences. By John Allen Paulos. Hill & Wang, New York, 1989, 135 pp. ISBN 0-8090-7447-8, $16.95.

LISA J. EVERED
Department of Mathematics, Iona College, New Rochelle, NY 10801

There is an old adage they tell to cub reporters: Never give the circus a bad review. The same underlying principle provides a cautionary motive for this reviewer. What reader of this journal, including this reviewer, could possibly do other than applaud Professor Paulos for giving us a book on such an important and timely subject? Nonetheless, having offered my applause I feel obligated to state that I do not think that this book lives up to the expectations that the title creates. I surely agree with Professor Paulos that there is a problem of innumeracy but I do not know the audience to whom this book is addressed, I do not think that the author has a clear understanding of the problem he is discussing, and I do not understand how he thinks that this book can help with the problem of innumeracy.

Surely all of the readers of this journal will agree with Professor Paulos that there is a problem of innumeracy and that, roughly, innumeracy refers to numbers and arithmetical manipulations in the same kind of way that illiteracy refers to letters and reading. The problem of innumeracy, then, is that there seem to be many people who are unable to cope with numbers, in a manner quite strikingly analogous to the way in which there are quite a few people who are unable to read.

If this be temporarily granted, subject to subsequent discussion, then we may begin by outlining the approach taken by Professor Paulos in his book. His first chapter, "Examples and Principles," is devoted to examples of arithmetical calculations involving, often, large numbers. He offers a one-paragraph explanation of scientific notation and then tells us that human hair grows at 10^{-8} miles per hour and that 5×10^{11} cigarettes are smoked each year in the United States. He calculates the total volume of human blood in the world and concludes that if walls were put around Central Park in New York City then all the blood in the world would cover the park to a depth of about twenty feet. He cites Archimedes and approximates his calculation of the total number of grains of sand required to fill up the earth and the heavens. Next he calculates some combinations, including the number of five-card hands dealt at poker. Then he estimates the probability of acquiring AIDS heterosexually. Finally, he calculates the probability that anyone who takes a deep breath inhales a molecule exhaled by Caeser in his dying breath.

Chapter 2, "Probability and Coincidence," gives a number of probabilistic calculations. We are told that "A tendency to drastically underestimate the frequency of coincidences is a prime characteristic of innumerates. . . ." After some

examples, Professor Paulos proceeds to the calculations. Anyone who teaches a course in basic statistics will be familiar with many of the examples: the birthday problem, the game of chuck-a-luck, the 1964 mugging in Los Angeles involving a blonde woman with a ponytail, and Pete Rose's National League record of hitting in 44 consecutive games. Chapter 3, "Pseudoscience," briefly discusses biorhythms, parapsychology, dreams, astrology, numerology, and UFO's. These subjects are, at best, only marginally related to innumeracy and one might well wonder why they are included. The author does, however, include a section on conditional probabilities in this chapter.

Chapter 4, "Whence Innumeracy," presents the author's ideas concerning the basic question he raises: "Why is innumeracy so widespread even among otherwise educated people?" His answer, in summary form, is: "The reasons, to be a little simplistic, are poor education, psychological blocks, and romantic misconceptions about the nature of mathematics." The author's comments here are both interesting and insightful. "Romantic misconceptions" include the belief that a concern for numbers is antithetical to an appreciation of the grand questions, that mathematics is low-level and technical, and that it is limiting and opposed to human freedom. In this chapter the author also includes a digression on the use of a logarithmic safety index for evaluating risks. Chapter 5, "Statistics, Trade-offs, and Society," is the final chapter. It is pretty much of a grab-bag: voting paradoxes, the prisoner's dilemma, death days, Type I and Type II errors, polling, correlation, and other topics. The author claims that most of these examples share the general problem of the balancing of conflicting concerns, but for me this common denominator was not much in evidence.

With this short summary in hand, let us return to the questions asked in the first paragraph of this review. To whom is this book addressed? It is fairly obvious that no innumerate person is going to read it. In fact, probably only quite highly numerate people are likely to find the book interesting. If the book is viewed as a compendium of often counterintuitive, attention-gaining numerical examples for teachers, then it can be given high praise since Professor Paulos has missed very few such examples. But the author's telegraphic style and encapsulated discussions will be incomprehensible to any "innumerate" who happens to pick up the book.

More generally, who is interested in the kinds of numerical calculations the author exemplifies in his Chapter 1? Archimedes calculated the number of grains of sand it would take to fill the universe because he wanted to show that such a calculation could be done. Granted the Greek system of numeration, it was no mean feat. But apart from the transient sense of triumph, did anyone really care what number he came up with? The same thing is even more obviously true with Professor Paulos' own example: the depth of all of the human blood in the world if it were stored in a reservoir based in Central Park. Many numerates would have to agree with the innumerates: Who cares? Nonetheless, some of us definitely enjoy working such things out. But the key phrase is "some of us." In fact, it is very few of us, and most mathematicians are not included. My observation is that the enjoyment of such calculations is not a common characteristic, even among numerate people. The reason for the enjoyment seems to be partly that of the mountaineer: It is there! But the other part may be more a matter of humor than of anything else—and Professor Paulos is the author of a well-known book on mathematics and humor. Thus, if someone asks whether a ballpoint pen costs more per mile than

does a typical car there is, for some of us, a certain entertainment value to the question and, hence, to the answer. Nonetheless, all of this suggests that in the present book Professor Paulos has undertaken the easy task of preaching to the already converted. The hard task is converting the unbelievers.

More importantly, it is folly and false to assume that there is any relationship between these kinds of calculations and the problem of innumeracy. It is false for the reasons already noted, namely, that highly numerate people do not find such calculations of any interest. It is folly because it misconstrues the problem of innumeracy. The analogy with illiteracy is helpful here. In the area of reading, the social problem is generally recognized to be that of what is called functional illiteracy. This includes all of those persons who are unable to read signs and the most simple instructions. We can all agree wholeheartedly that such persons do represent a social problem because they are undergoing hazards that most of us avoid, and because they are unable to participate in most socially desirable activities. However, we surely do not argue that anyone has some degree of illiteracy that is of social concern simply because that person does not ever read Shakespeare or Thomas Mann. Some of us may privately believe that life without reading Shakespeare is life misspent but we do not establish scales of illiteracy based on this belief. This, it seems to me, is what Professor Paulos is doing when he equates, at least by implication, innumeracy with the inability to perform or follow the relatively intricate calculations he presents.

Another pervasive feature of the author's approach can be illustrated by one of his probability examples, the birthday problem. There are many informed people who argue that one of the most important subjects that ordinary individuals, including innumerates, would find useful to know is elementary probability theory. Unfortunately, as any teacher of beginning statistics can testify, many highly intelligent students find the subject distressingly difficult. One reason for this is undoubtedly that the methods of solution of elementary probability problems often seem to the student to smack of ad hoc conjuring tricks. There are not at all the kind of stable procedures that the student finds, say, in algebra or differential calculus. This fact ought to go far towards defining what a book like this one ought to say about probability theory. It seems to me that Professor Paulos' approach is quite otherwise. Consider, then, the birthday problem. It is, of course, quite counterintuitive that for a randomly selected group of 23 persons the probability is a little more than one-half that at least two will have the same day of the year for their birthday. But how does working the argument out benefit an innumerate? Does knowing the answer, or the argument, help the reader to the slightest degree in estimating or calculating the answer to any variation of the problem? Absolutely not. Let the reader of this review try these variations as examples. There are 12,000 taxicabs in New York City. How many riders are required to make the probability one-half that the same taxicab has been used at least twice? There are about 2.6 million five-card poker hands. How many hands must be dealt to make the probability one-half that at least one hand has occurred twice? Return to the birthday problem itself. How many persons must be randomly selected so that the probability is one-half that at least three have the same birthday? At least five? At least ten? How many persons so that the probability is .9 that at least two have the same birthday? (The answers are at the end of this review.) I think the reader will be convinced that neither the answer nor the method of the birthday problem is easily

generalizable. What, then, is the advantage of presenting the argument to an innumerate reader?

My hypothesis is that Professor Paulos has been misled by his own definition of the problem of innumeracy. By defining numeracy in terms of the ability to duplicate the solutions of his examples he has put it beyond the reach of most innumerates. In fact, I suspect that almost all of the readers of his book will be persons who already share the author's opinions and his enjoyment of calculation for its own sake. An innumerate person who chances to pick up this book will probably recoil in fear and trembling. It is most unlikely that such a person would be able to negotiate the arithmetical difficulties presented by the author.

With a more nearly correct definition of the problem of innumeracy there is, I believe, some hope for its solution. We need to start with a definition of functional innumeracy. This would surely mean such things as inability to calculate an 8.25% sales tax on a purchase, inability to calculate the price that results from a 20% reduction in price, inability to fill out the income tax forms, inability to calculate the paint required for an apartment given the square feet covered per gallon, and so forth. These are the kinds of things that everyone needs to be able to do—and can learn to do. Whether these things can be self-taught from a book is not clear and is perhaps doubtful. But it is certainly clear that these are desirable and achievable goals.

The problem of finding the minimal kind of knowledge of probabilities that is needed to escape functional innumeracy is more vexing. Perhaps the most important kinds of knowledge have nothing to do with probabilistic calculations. There is a fair amount of evidence that living in accordance with rational prescriptions requires ability to estimate probabilities based on one's own knowledge ("probability assessing," in the literature), an ability to behave rationally regarding risks, and an awareness of the general overestimation of small probabilities (the "over-under effect," in the literature). There is a considerable body of experimental evidence that these are all learnable skills. Something clearly needs to be done in this area in any general assault on the problem of innumeracy.

As already mentioned, it is not clear whether any or all of this might be accomplished through self-study from an appropriate book. But it is surely worthy of note that Professor Paulos' section on a logarithmic safety index is a very good initial statement on how to think about risks. One can easily imagine that if such a book can be successful then perhaps Professor Paulos is precisely the person to write it.

(Answers to questions: 128 rides; 1900 poker hands or nine hours of play for seven players at the rate of two minutes per hand; 87 persons for $\geqslant 3$; 310 for $\geqslant 5$; 1170 for $\geqslant 10$; 41 persons for a probability of .9.)

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S(13-16). *Higher Mathematics in Problems and Exercises.* P.E. Danko, A.G. Popov, T. Ya. Kozhevnikova. MIR (US Distr: Imported Pub), 1988, $14.95 each. *Part 1,* 400 pp, [ISBN: 0-8285-3957-X]; *Part 2,* 480 pp. [ISBN: 0-8285-3958-8] The Russian equivalent of the Schaum Outline Series: a collection of specially selected exercises and solutions aimed at revealing the essential features of the main concepts and theorems of undergraduate mathematics (especially engineering mathematics). Each section begins with a brief summary of relevant theoretical underpinnings, then proceeds to sample problems and solutions, and concludes with exercises for the student; numerical answers in the back. *Part 1* covers analytic geometry (with elements of vector algebra), linear algebra, calculus, elementary analysis, linear programming. *Part 2* includes multivariable calculus, differential equations, probability theory, complex variables, operational calculus, numerical analysis, calculus of variations. The two volumes contain over 2800 examples and exercises; you can't beat the price! Especially appropriate for the college bookstore. LCL

Education, S, P. *Mathematics Projects Handbook, Third Edition.* Adrien L. Hess, Glenn D. Allinger, Lyle E. Andersen. NCTM, 1989, v + 52 pp, $5 (P). [ISBN: 0-87353-283-X] Designed as a guide for teachers and students in choosing mathematics projects, from classroom learning experiences to exhibits for entrance in competitions. This edition contains student guides and up-dated references, and places more emphasis on topics such as discrete mathematics, probability, and statistics (*First Edition,* TR, January 1978; *Second Edition,* TR, October 1982). SB

Education, P, L. *Results from the Fourth Mathematics Assessment of the National Assessment of Educational Progress.* Ed: Mary Montgomery Lindquist. NCTM, 1989, viii + 173 pp, $15 (P). [ISBN: 0-87353-274-0] Examples of test items from the 1986 assessment for objectives in seven content areas: mathematical methods, discrete mathematics, data analysis, measurement, geometry, relations and functions, and numbers and operations. Summarizes and interprets results on individual items for grades 3, 7, and 11. Analyzes performance differences among minority and gender groups. More detail than has appeared in NCTM journals. MW

Education, S(16-17). *How to Use Conjecturing and Microcomputers to Teach Geometry.* Daniel Chazan, Richard Houde. NCTM, 1989, vii + 56 pp, $5 (P). [ISBN: 0-87353-279-1] Answers questions posed by teachers over five years of workshops on using *Geometric Supposer* software to teach geometry and engage students in behavior "like working members of the community of mathematicians." Detailed suggestions for planning, teaching, and testing. Excellent for teachers planning to use the *Supposers;* too specific for transfer to other situations. MW

History, S, P. *Charles Loewner, Collected Papers.* Ed: Lipman Bers. Birkhauser Boston, 1988, xii + 517 pp, $79. [ISBN: 0-8176-3377-4] The published papers of Charles Loewner. His contributions in the areas of complex analysis and differential geometry are part of the classical literature in these subjects. LCL

Logic, T(15-16: 1), L. *Notes on Logic and Set Theory.* P.T. Johnstone. Cambridge U Pr, 1987, x + 111 pp, $34.50; $12.95 (P). [ISBN: 0-521-33502-7; 0-521-33692-9] Text aims to present "everything that the well-educated mathematician in the late twentieth century needs to know about the foundations of his subject." Covers first-order logic through completeness theorem, recursion theory, Zermelo-Fraenkel axioms, ordinal and cardinal arithmetic, incompleteness theorems. Assumes knowledge of abstract algebra. KES

Logic, P. *Cartesian Closed Categories of Domains.* A. Jung. CWI Tract, V. 66. Mathematisch Centrum, 1989, 110 pp, Dfl. 17,80 (P). [ISBN: 90-6196-376-1] A study of objects which have no finite description and functions defined on them. The objective is to strike a balance between the idealistic and the realistic; the framework for satisfying these requirements is that of algebraic directed-complete partial orders (called domains). This tract presents the author's

thesis results. LCL

Logic, S(16-18), P. *Proofs and Types.* Jean-Yves Girard. Transl: Paul Taylor, Yves Lafont. Tracts in Theoret. Comput. Sci. Cambridge University Pr, 1989, xi + 176 pp, $34.50. [ISBN: 0-521-37181-3] A short course on typed λ-calculus (Prawitz' natural deduction, sequent calculus, Gödel's System T, System F) given at the Université Paris VII in 1986-87. LCL

Foundations, P. *Toposes and Local Set Theories, An Introduction.* J.L. Bell. Logic Guides, V. 14. Clarendon Pr, 1988, xii + 267 pp, $75. [ISBN: 0-19-853274-1] Toposes (categories with features—products, subobjects, terminal object, power objects—which enable them to be considered generalizations of the category of sets), and local set theories (where types replace sets as primitives) serve as alternatives to ordinary set theory as foundation for mathematics, yielding intuitionistic logic of types. Toposes are presented as models of theories, "domains of discourse" in which mathematical concepts may be defined, interpreted. RM

Combinatorics, P. *q-Series and Partitions.* Ed: Dennis Stanton. IMA, V. 18. Springer-Verlag, 1989, xi + 212 pp, $24. [ISBN: 0-387-97086-X] Proceedings of a workshop in q-series and partitions held at the IMA, Minneapolis, March 1988. Includes papers on identities, unimodality of Gaussian polynomials, constant term problems, and orthogonal polynomials. LC

Number Theory, P. *Lecture Notes in Mathematics-1380: Number Theory.* Eds: H.P. Schlickewei, E. Wirsing. Springer-Verlag, 1989, v + 266 pp, $24.30 (P). [ISBN: 0-387-51397-3] Proceedings from the 15th Journées Arithmétiques held at the University of Ulm in 1987. Contains only papers not appearing elsewhere ranging over many areas of current research. Included in the paper of G. Frey is progress on a promising approach to proving Fermat's Last Theorem. GG

Algebra, T(14-15). *The Number Systems: Foundations of Algebra and Analysis, Second Edition.* Solomon Feferman. Chelsea, 1989, xii + 418 pp, $23.95. [ISBN: 0-8284-0333-3] A course in the foundation of the number system. The author begins with elementary set theory, assumes the existence of the positive integers, and then develops its properties and properties of the integers, rationals, the reals (via cuts), and the complex numbers. Also discussed are integral domains, polynomials, and extension fields. Worthy of consideration as a text for the "transitional" course for mathematics majors. A reprinting of the 1964 edition. SG

Algebra, L. *A Survey of Lie Groups and Lie Algebras with Applications and Computational Methods.* Johan G.F. Belinfante, Bernard Kolman. Classics in Appl. Math., V. 2. SIAM, 1989, iii + 164 pp, $24.50 (P). [ISBN: 0-89871-243-2] Corrections of minor misprints, but no major changes from the 1972 original edition (TR, January 1973). LC

Algebra, P, L. *Lexikon der Algebra.* Günther Eisenreich. Akademie-Verlag, 1989, 677 pp, 80 DM. [ISBN: 3-05-500231-8] A modern, carefully organized and extensive German dictionary of algebra. Statements of important theorems; many references to literature. JD-B

Algebra, T(15-16: 1). *Algebra for Computer Science.* Lars Gårding, Torbjörn Tambour. Universitext. Springer-Verlag, 1988, ix + 198 pp, $29.80 (P). [ISBN: 0-387-96780-X] Very concise presentation of number theory and abstract algebra and their applications to computer science. Proofs of many theorems left as exercises. Assumes mathematical maturity and elementary linear algebra. KES

Algebra, P. *Galois Groups over Q.* Eds: Y. Ihara, K. Ribet, J.-P. Serre. Math. Sci. Res. Inst., V. 16. Springer-Verlag, 1989, x + 449 pp, $45. [ISBN: 0-387-97031-2] Inspired by a week-long conference at the Berkeley Mathematical Science Research Institute, March 23-27, 1987. Principal themes of the conference are extensions of Q with finite simple Galois groups; Galois actions on fundamental groups, nilpotent extensions of Q arising from Fermat curves, and the interplay between Gauss sums and cyclotomic units; and representations of Gal $(\overline{Q}/Q)$ with values in $GL(2)$; deformations and connections with modular forms. SB

Algebra, T*(18: 1-4), S, P, L. *Basic Algebra II, Second Edition.* Nathan Jacobson. WH Freeman, 1989, xvii + 686 pp, $47.95. [ISBN: 0-7167-1933-9] A number of changes occur in the segment of the book devoted to commutative algebra, plus a number of other improvements and corrections, in addition to some new exercises and an updated index. (1980 edition, TR, January 1981.) LCL

Calculus, T(13: 1). *Applied Calculus: A First Course, Second Edition.* Shirley O. Hockett, Martin Sternstein. Robert E Krieger, 1989, x + 654 pp, $44.50. [ISBN: 0-89464-351-7] Intended for business, biology, and social science students. No trigonometric functions. Lots of applied examples and exercises, generally routine and artificial. Brief sections on separable first order differential equations, multivariable functions, and Lagrange multipliers. "Things to Remember" sections and extra exercises added for this edition. (1979 Van-Nostrand Reinhold edition, TR, June-July 1979.) JO

Calculus, T*(13-14: 1-4), S*, L. *Introductory Mathematics Through Science Applications.* John Berry, Allan Norcliffe, Stephen Humble. Cambridge University Pr, 1989, xii + 547 pp, $79.50; $29.95 (P). [ISBN: 0-521-28446-5; 0-521-24119-7] An applications-based introduction to the first two years of college mathematics. Designed to "introduce the mathematics from a physical point of view" without overwhelming the student with mathematical rigor. Stressing the application of calculus to the sciences, each chapter begins with several examples "setting the new techniques to be introduced in the context of the physical world," and includes two more "worked examples" sections and an extensive exercise set (with answers in the Appendix). First

five chapters provide a nice introduction to functions and their uses, then three chapters each on differentiation and integration. Introduces numerical methods in context rather than as separate topics. Also includes substantial sections on ordinary and partial differential equations, probability, and statistics. Could be a good source for problems or as a text (note paperback option). Give it a look. SB

Real Analysis, P. *Regular Variation.* N.H. Bingham, C.M. Goldie, J.L. Teugels. Encyclop. of Math. & Its Applic., V. 27. Cambridge University Pr, 1989, xix + 494 pp, $34.50 (P). [ISBN: 0-521-37943-1] Changes from the 1987 hardback edition (TR, May 1988) include recent progress in the field (e.g., R. Trautner's covering principle) and an expanded reference section. MLR

Complex Analysis, T(17: 1), S, L. *Elliptic Functions and Applications.* Derek F. Lawden. Appl. Math. Sci., V. 80. Springer-Verlag, 1989, xiv + 334 pp, $59. [ISBN: 0-387-96965-9] An introductory book intended by the author to serve as a text for (British) undergraduates and "practicing applied mathematicians and engineers." Very readable anytime after a first course in complex analysis. Focuses on geometrical and physical, as opposed to number-theoretical aspects. Includes numerous exercises and several tables of computations. GG

Partial Differential Equations, P. *Heat Kernels and Spectral Theory.* E.B. Davies. Tracts in Math., V. 92. Cambridge U Pr, 1989, ix + 197 pp, $49.50. [ISBN: 0-521-36136-2] A concise, yet comprehensive, study of a classical subject—second-order linear elliptic differential equations—with special emphasis on modern techniques arising from theories of quadratic forms and Sobolev spaces. Such techniques lead to sharp, computable bounds on the integral kernels in question. Most results are obtained for regions in Euclidean space, but a final chapter covers operators on Riemannian manifolds. PZ

Operator Theory, S(18), P. *Variational Principles for Nonpotential Operators.* V.M. Filippov. Transl. of Math. Mono., V. 77. AMS, 1989, xiii + 239 pp, $99. [ISBN: 0-8218-4529-2] Written to fill a perceived need for a monograph on variational principles for equations with nonsymmetric operators. Chapters I and II develop the theoretical foundations; Chapter III treats variational principles for linear boundary value problems; Chapter IV concerns nonlinear equations. Chapter commentaries; bibliography. JS

Functional Analysis, P. *Lecture Notes in Mathematics-1382: Metaplectic Groups and Segal Algebras.* Hans Reiter. Springer-Verlag, 1989, xi + 128 pp, $13.50 (P). [ISBN: 0-387-51417-1] Explores a connection between Weil's metaplectic group and a certain Segal algebra. MLR

Functional Analysis, P. *Characterizations of Banach Spaces Not Containing ℓ^1.* D. van Dulst. CWI Tracts, V. 59. Mathematisch Centrum, 1989, 163 pp, Dfl. 25.30 (P). [ISBN: 90-6196-366-4] Covers the problem of determining whether or not a Banach space contains an isomorphic copy of ℓ^1. LC

Analysis, P. *Commutative Harmonic Analysis.* Ed: David Colella. Contemp. Math., V. 91. AMS, 1989, xi + 305 pp, $36 (P). Proceedings of an international conference on harmonic analysis held in July 1987 at St. Lawrence University, sponsored by GTE. Volume includes twenty-three papers on a wide range of topics. PZ

Analysis, T(18), P. *Orthogonal Series.* B.S. Kashin, A.A. Saakyan. Transl. of Math. Mono., V. 75. AMS, 1989, xii + 451 pp, $146. [ISBN: 0-8218-4527-6] Presents the main idea and methods of the theory of general orthogonal series. Includes chapters on the Haar system, trigonometric and Walsh systems, the Faber-Schauder and Franklin systems, and results on convergence, divergence, and representations of orthogonal series. Aimed at the beginning mathematics student with some background in functional analysis and complex analysis. LC

Algebraic Geometry, T(16-18: 1), S, P, L. *Introduction to Algebraic Curves.* Phillip A. Griffiths. Transl. of Math. Mono., V. 76. AMS, 1989, x + 221 pp, $93. [ISBN: 0-8218-4530-6] Translation from the Chinese text of notes based on a course taught by the author at Beijing University in 1982. Chapters include fundamental concepts (curves in the complex projective plane, Riemann surfaces, meromorphic functions, differential forms, complex manifolds, algebraic varieties), the Normalization Theorem, the Riemann-Roch Theorem, and Abel's theorem. Good first reference for non-specialists; nice use of graphics; numerous exercises within the text; annotated bibliography, and test problems in the Appendix. Assumes working knowledge of elementary complex function theory and algebra. SB

Algebraic Geometry, P. *Regular Differential Forms.* Ernst Kunz, Rolf Waldi. Contemp. Math., V. 79. AMS, 1988, ix + 153 pp, $20 (P). [ISBN: 0-8218-5085-7] Studies, in a general algebraic setting, local properties of sheaves of regular differential forms on a quotient of Noetherian schemes. This work extends to the "relative" case a variety of results already known in the "absolute" case, i.e., for varieties over fields. PZ

Algebraic Geometry, S(18), P. *Torsion Theoretic Algebraic Geometry.* Frederick W. Call. Papers in Pure & Appl. Math., No. 82. Queen's U, 1989, 118 pp, (P). Brief lecture notes which could reasonably supplement an algebraic geometry course (categories, schemes, cohomology, etc.). GG

Algebraic Geometry, P. *Lecture Notes in Mathematics-1389: Algebraic Curves and Projective Geometry.* Eds: E. Ballico, C. Ciliberto. Springer-Verlag, 1989, 288 pp, $24.30 (P). [ISBN: 0-387-51509-7] Proceedings of a 1988 conference held in Trento, Italy containing over twenty papers on the topic. SG

Differential Geometry, P. *Lecture Notes in Mathematics-1377: Singularity Theory, Rod Theory, and Symmetry-Breaking Loads.* John F. Pierce. Springer-Verlag, 1989, iv + 177 pp, $16.20 (P). [ISBN: 0-387-51304-3] Uses methods of modern

analysis and singularity theory to examine how orbits for rods subject to symmetric loads and in equilibrium change when the loads are perturbed so that the symmetry is broken. For applied mathematicians and researchers in mechanics. Poses new problems. DFA

Geometry, T*(15-16: 1), L*. *A Course in Modern Geometries.* Judith N. Cederberg. Undergrad. Texts in Math. Springer-Verlag, 1989, xii + 232 pp, $42. [ISBN: 0-387-96922-5] For junior-senior level survey course in geometry. Seeks to develop geometrical insights into abstract-appearing results. Brief history of Euclid's *Elements* with major emphasis on geometries invented after Euclid. Synthetic approach followed by analytical treatment in later chapters. Axiomatics, finite geometries, transformations, and an introduction to projective geometry. Excellent reading and viewing lists. Exercises, many of which complement or extend the text. Prerequisites are some high school geometry and a bit of background in linear algebra. Nicely paced. JDEK

Geometry, S*(16), P*, L*. *The Ethics of Geometry: A Genealogy of Modernity.* David Rapport Lachterman. Routledge, Chapman & Hall, 1989, xiv + 255 pp, $17.95 (P); $49.50. [ISBN: 0-415-90141-3; 0-415-90053-0] Scholarly treatment of the relationship between philosophy and mathematics. "Essential differences in the *ethos* of mathematics, i.e., the customary ways of undertaking and understanding mathematical procedures and their objects, provide insight into the fundamental issues in the quarrel of moderns with ancients." Compares the roles of constructions in Euclid's geometry and in that of Descartes. Extensive notes with references to both primary and secondary sources. Invites comparison with Fowler's *The Mathematics of Plato's Academy* and Knorr's *The Ancient Tradition of Geometric Problems.* Heavy going. JDEK

Geometry, S(18), L. *Global Differential Geometry.* Ed: S.S. Chern. Stud. in Math., V. 27. MAA, 1989, ii + 354 pp, $36.50. [ISBN: 0-88385-129-6] Updated version of V. 4 of same series. Ten essays on fundamental topics, including four new chapters on vector bundles with connection, minimal surfaces, Riemannian comparison constructions, and manifolds of non-positive curvature. JO

Geometry, P. *Lecture Notes in Mathematics-1355: Computational Synthetic Geometry.* Jürgen Bokowski, Bernd Sturmfels. Springer-Verlag, 1989, 168 pp, $16.20 (P). [ISBN: 0-387-50478-8] Discusses algorithmic aspects of certain realizability problems in discrete geometry. Problems covered include those of constructing polytopes from simplicial complexes, vector geometries from incidence structures, and hyperplane arrangements from oriented matroids. LC

Algebraic Topology, P. *Algebraic Topology.* Eds: Mark Mahowald, Stewart Priddy. Contemp. Math., V. 96. AMS, 1989, xiii + 350 pp, $41 (P). [ISBN: 0-8218-5102-0] Proceedings of the International Conference on Algebraic Topology held at Northwestern University, Evanston, Illinois from March 21-24, 1988. Contains papers either presented or solicited at the conference. SB

Algebraic Topology, P. *Groups Acting on Graphs.* Warren Dicks, M.J. Dunwoody. Stud. in Adv. Math., V. 17. Cambridge U Pr, 1989, xvi + 283 pp, $69.50. [ISBN: 0-521-23033-0] Research monograph on groups acting on low-dimensional topological spaces. Includes standard results on groups acting on trees, Boolean rings of graphs, characterization of Poincaré duality groups, and the equivalent loop and sphere theorems. Assumes familiarity of homological algebra and algebraic topology. LC

Dynamical Systems, S(15-17), P*, L.** *Chaos and Fractals: The Mathematics Behind the Computer Graphics.* Eds: Robert L. Devaney, Linda Keen. Proc. of Symposia in Appl. Math., V. 39. AMS, 1989, ix + 148 pp, $31. [ISBN: 0-8218-0137-6] Expanded lectures from the August 1988 AMS Short Course (at the Providence Centennial meeting): Smale's horseshoe map, chaotic attractors, Julia and Mandelbrot sets, fractals, and applications of iterated functions to data compression. Provides substantive introductions to many related topics revealing the mathematics behind the glitz. Many color plates enhance the appeal of this attractive volume. LAS

Dynamical Systems, S(17-18), P. *Chaos in Dissipativen Systemen.* Ronald W. Leven, Bernd-Peter Koch, Bernd Pompe. Friedr. Vieweg & Sohn, 1989, 189 pp, DM 32 (P). [ISBN: 3-528-06356-4]

Probability, T(17: 2), S. *The Theory of Probability.* B.V. Gnedenko. MIR (US Distr: Imported Pub), 1988, 392 pp, $12.95. Translated from Russian. Covers the fundamentals of probability theory including the concept of probability, independent trials, random variables and distribution functions, expectation, limit theorems, laws of large numbers, characteristic functions, stochastic processes, and queueing theory. Good use of graphs; teaches through examples; emphasizes limit theorems; does not cover combinatorics; notation complicated; organization of text hard to follow for undergraduates. Good theoretical resource text. (1967 Chelsea edition, TR, February 1968; 1976 MIR edition, TR, August-Spetember 1976; 1978 MIR edition, TR, November 1981.) JJ

Probability, S(17-18). *Generalized Poisson Distributions: Properties and Applications.* P.C. Consul. Stat.: Textbooks & Mono., V. 99. Marcel Dekker, 1989, x + 302 pp, $89.75. [ISBN: 0-8247-7863-4] Complete treatment of the generalized Poisson distribution and generalized Poisson process. Covers both applications and theoretical results. Develops confidence intervals and statistical tests for the two parameters, λ and θ, using maximum likelihood estimation. Includes computer programs for calculating probabilities, confidence intervals, etc. Excellent reference text. JJ

Elementary Statistics, T(13-14: 1). *Principles of Statistics.* Paul A. Herzberg. Robert E Krieger,

1989, xvi + 513 pp, $48.50, [ISBN: 0-89464-374-6]; *Study Guide,* Elke U. Weber, 343 pp, $16.50 (P). [ISBN: 0-89464-409-2] Introductory text emphasizing concepts and requiring only algebra. Large portion devoted to hypothesis testing, including tests for categorical and ranked variables. Careful attention given to requirements for procedures to be valid. Discusses elementary probability, descriptive statistics, several measures of association, regression, and ANOVA. *Study Guide* contains objectives, tips and reminders, and ten multiple-choice questions for each chapter of text. RWJ

Elementary Statistics, S*, P, L.** *Scientific Reasoning: The Bayesian Approach.* Colin Howson, Peter Urbach. Open Court, 1989, xiii + 312 pp, $16.95 (P); $34.95. [ISBN: 0-8126-9085-0; 0-8126-9084-2] A comprehensive *apologia* for Bayesian methods as providing a firmer logical foundation for scientific inference than the more fashionable classical theories of inference (hypothesis testing, randomized trials, confidence intervals). Explains and defends Bayesian methods—especially against the charge of subjectivity—and also criticizes classical statistics as inconsistent, illogical, and contrary to the actual practice of science. Clearly written; carefully referenced; accessible to non-specialists. LAS

Computational Statistics, T(13-14: 1). *Guide to Mathematical Modelling.* Dilwyn Edwards, Mike Hamson. Math. Guides. Macmillan Education (US Distr: Scholium Intern), 1989, x + 277 pp, $23.50 (P). [ISBN: 0-333-45935-0] A first-year course in undergraduate mathematics, statistics, or computing. Provides an approach to formulate and solve real problems in terms of mathematics and statistics. Modelling problems presented include those involving random features which demand statistical analysis, and situations which give rise to differential equations. Communication skills addressed in one full chapter, and ample interesting example models are presented in another. Many problems posed throughout. MSS

Statistics, P. *Lecture Notes in Statistics-57: Statistical Modelling.* Eds: A. Decarli, *et al.* Springer-Verlag, 1989, ix + 343 pp, $37.80 (P). [ISBN: 0-387-97097-5] Proceedings of the July 1989 joint meeting of the GLIM 89 Conference and the Fourth International Workshop on Statistical Modelling held in Trento, Italy. Most of the thirty-seven papers deal with the use and/or development of GLIM software for general linear models. RWJ

Statistics, S(17-18). *Transformation and Weighting in Regression.* Raymond J. Carroll, David Ruppert. Mono. on Stat. & Appl. Prob. Chapman & Hall, 1988, x + 249 pp, $45. [ISBN: 0-412-01421-1] Examines regression methods for data in which the classical regression assumptions of constant variance and normally distributed data are violated. Methods discussed fit into two groups: weighting and weighted least squares, and data transformations. Discusses maximum likelihood estimators and transformations on both sides of regression equation. Many examples with data; good use of graphs. JJ

Programming, P. *A Guidebook to Fortran on Supercomputers.* John M. Levesque, Joel W. Williamson. Academic Pr, 1989, vii + 218 pp, $39.95. [ISBN: 0-12-444760-0] Guide to developing and restructuring Fortran programs for supercomputers (those that exploit parallelism) so that optimizing compilers can translate code to take full advantage of the underlying architecture. Discussion of architecture and vectorization techniques; includes many restructuring examples. RM

Computer Systems, S(15-16), P, L. *Supercomputers and their Use, Revised Edition.* Christopher Lazou. Clarendon Pr, 1988, xi + 263 pp, $55; $27.50 (P). [ISBN: 0-19-853815-4; 0-19-853759-X] Relatively accessible introduction to and overview of supercomputers: architecture and organization, underlying technology, performance, illustrative applications, likely future developments (some now dated), discussions of novel architectures. (Revision of 1986 edition.) RM

Theory of Computation, P. *Lecture Notes in Computer Science-379: Mathematical Foundations of Computer Science 1989.* Eds: A. Kreczmar, G. Mirkowska. Springer-Verlag, 1989, viii + 605 pp, $45.50 (P). [ISBN: 0-387-51486-4] Forty research papers originally presented at the Fourteenth Symposium in this series held in Poland in 1989. The papers address advanced and theoretical research in such subareas of computer science as automata theory, logic programming, theory of concurrent processes, and formal connectness proofs. GMS

Theory of Computation, P. *Theory of Deductive Systems and Its Applications.* S. Yu. Maslov. Transl: Michael Gelfond, Vladimir Lifschitz. Ser. in Foundat. of Comput. MIT Pr, 1987, x + 151 pp, $20. [ISBN: 0-262-13223-0] A deductive system consists of initial objects and rules for generating new objects. This monograph presents mathematics of deductive systems based on Post's canonical calculi and describes applications in economics, biology, computer science, and operations research. Translation of 1986 Russian publication. KES

Theory of Computation, P, L. *Computational Complexity Theory.* Ed: Juris Hartmanis. Proc. of Symp. in Appl. Math., V. 38. AMS, 1989, ix + 128 pp, $33. [ISBN: 0-8218-0131-7] Introductory survey lectures from the Short Course at the January 1988 AMS meeting in Atlanta. Six papers for the nonspecialist, including an overview of the subject by the editor. The others concern the isomorphism conjecture and sparse sets; restricted relativizations of complexity classes; descriptive and computational complexity (concerns parallel processing); complexity issues in cryptography; interactive proof systems. DFA

Theory of Computation, P. *Resolution of Equations in Algebraic Structures, Volume 2: Rewriting Techniques.* Eds: Hassan Aït-Kaci, Maurice Nivat. Academic Pr, 1989, xx + 379 pp, $64.95. [ISBN: 0-

12-046371-7] The second of two volumes containing papers originally presented at a colloquium held in Austin, Texas, May 1987. The twelve papers in this monograph are all directed at presenting advanced research in using the resolution method for solving symbolic equations. GMS

Artificial Intelligence, P. *Automated Theorem-Proving in Non-Classical Logics.* Paul B. Thistlewaite, Michael A. McRobbie, Robert K. Meyer. Wiley, 1988, 154 pp, $19.95 (P). [ISBN: 0-273-08781-9] Discusses history of and motivations for nonclassical theorem-proving, and describes development of program for deciding theoremhood in relevant logic. KES

Artificial Intelligence, P. *Blackboard Architectures and Applications.* Eds: V. Jagannathan, Rajendra Dodhiawala, Lawrence S. Baum. Perspect. in Artif. Intellig., V. 3. Academic Pr, 1989, xxix + 524 pp, $39.95. [ISBN: 0-12-379940-6] Twenty-one papers from the AAAI-87 and -88 Blackboard workshops presenting recent developments. Topics: organization and control, advanced architectures—concurrency and parallelism, real-time systems, development environments, applications. An introduction traces the history of the blackboard concept. Over 250 references. DFA

Applications, S, P. *Urban Rank-Size Hierarchy: A Mathematical Interpretation.* James W. Fonseca. Mono. No. 8. Institute of Mathematical Geography (2790 Briarcliff, Ann Arbor, MI 48105), 1989, 86 pp, $15.95 (P). We know that the Fibonacci numbers, $F(n)$, appear "in nature" in certain growth patterns (e.g., conical shells, sunflowers, pine cones, galaxies). Here the author shows that in the United States, the urban rank-size hierarchy can be modeled quite accurately in this same way. One manifestation of this is that the ratio of the $F(n+1)^{st}$ ranked urban area of the $F(n)^{th}$ is approximately ϕ, the golden ratio. LCL

Applications (Biological Science), S(16-17), P, L. *Dynamic Modeling in Behavioral Ecology.* Marc Mangel, Colin W. Clark. Mono. in Behavior & Ecology. Princeton U Pr, 1988, xii + 308 pp, $15.95 (P); $45. [ISBN: 0-691-08506-4; 0-691-08505-6] Self-contained introduction to construction and use of dynamic stochastic optimization models in behavioral ecology. Based on evolutionary principles; approach allows consideration of many types of behavior; decisions to be considered simultaneously; incorporation of various constraints. Applications to hunting behavior, insect reproduction, migration of aquatic organisms, clutch size in birds, movement of spiders. RM

Applications (Economics), T(18: 2). *Advanced Econometric Methods.* Thomas B. Fomby, R. Carter Hill, Stanley R. Johnson. Springer-Verlag, 1984, xix + 624 pp, $35 (P). [ISBN: 0-387-96868-7] Paperback version of a comprehensive graduate-level text for econometrics. Covers most major topics in econometrics. Assumes readers who have had an econometric methods course, and who are familiar with matrix notation in regression. Econometric theory covered, but only proofs of basic theorems are given. Some data for analysis provided in exercises. (1984 hardcover edition, TR, February 1985.) JJ

Applications (Engineering), S(17-18), P. *Expert Systems for Engineering Design.* Ed: Michael D. Rychener. Academic Pr, 1988, xii + 306 pp, $39.95. [ISBN: 0-12-605110-0] Selection of papers on the application of expert systems techniques to engineering design problems. Discussion of existing systems, trends, with theme of synthesis, nature of expertise, integration of software, and graphics interfaces. RM

Applications (Fluid Dynamics), P. *The Linearization Method in Hydrodynamical Stability Theory.* V.I. Yudovich. Transl. of Math. Mono., V. 74. AMS, 1989, iii + 170 pp, $74. [ISBN: 0-8218-4528-4] Translation of the 1984 Russian edition. Applies Lyapunov's method of linearization to Navier-Stokes equations and the stability of steady and time-periodic fluid motion. Bibliography includes many references to the Soviet literature. JO

Applications (Physics), P. *Noise in Nonlinear Dynamical Systems, Volume 1: Theory of Continuous Fokker-Planck Systems.* Eds: Frank Moss, P.V.E. McClintock. Cambridge University Pr, 1989, xvi + 352 pp, $85. [ISBN: 0-521-35228-2] First of three volumes comprising a collection by different authors which may be read individually. This volume deals with the basic theory of stochastic nonlinear systems. It also includes a detailed history of the subject and the first English translation of the seminal 1933 paper of Pontriagin, Andronov, and Vitt. KS

Applications (Quality Control), T(17-18: 1), S. *Statistical Methods for Quality Improvement.* Thomas P. Ryan. Appl. Prob. & Appl. Stat. Wiley, 1989, xviii + 446 pp, $49.95. [ISBN: 0-471-84337-7] Graduate-level text and excellent resource on quality control. Provides extensive coverage of control charts including CUSUM charts, attribute control charts, and multivariate control charts. Basic introduction to probability and statistics covers distributions used in book. Other topics beyond control charts include graphical methods, linear regression, design of experiments (including factorials, fractional factorials, Taguchi methods), and analysis of means. Contains some examples and exercises, but could contain more examples presenting successful uses of these methods in actual companies. JJ

Reviewers

DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; LC: Laura Chihara, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; RWJ: Roger W. Johnson, Carleton; JJ: Jason Jones, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; JO: Jeff Ondich, St. Olaf; MLR: Margaret L. Reese, St. Olaf; KS: Karen Saxe, St. Olaf; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; KES: Kay E. Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MSS: Myriam S. Steinback, Macalester; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

MATHEMATICS: Queen and Servant of Science

by E. T. Bell

"Bell is a lively, stimulating writer, inoffensively crotchety and opinionated, with a good sense of historical circumstance, a fine impatience with humbug, a sound grasp of the entire mathematical scene, and a gift for clear and orderly explanation."

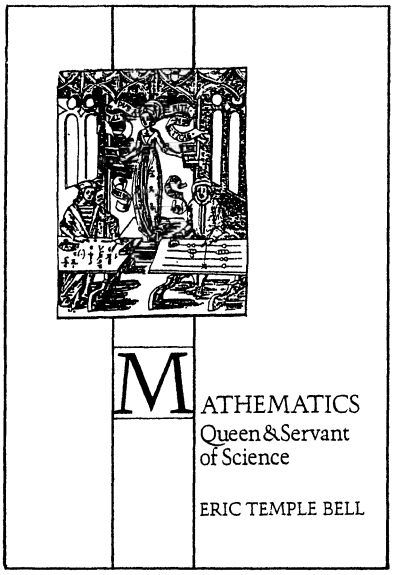
James R. Newman, *The World of Mathematics*

"Bell's writing style makes you feel as if you are a fortunate student spending time with a charming and gifted teacher. Not a history of mathematics, but a wide ranging exposition of ideas in the subject, the book is a classic and should be required reading for every teacher of mathematics."

Harold Jacobs, Grant High School, Van Nuys, California

"This book deserves a place in today's market. It is a much more popular work than most histories of the subject, and that is exactly what makes it accessible to students as well as to non-mathematicians. It is rewarding reading for . . . teachers and students of mathematics at all levels."

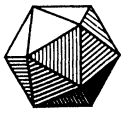
Morris Kline, Courant Institute



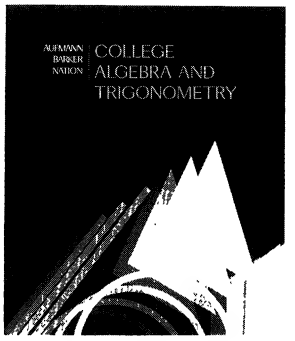
xv + 437 pp. Paperbound
Republished 1987.
ISBN-0-88385-446-3
List: $18.50
MAA Member: $14.00

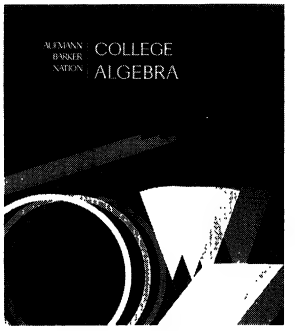
This is an absorbing account of the development in pure and applied mathematics from the geometry of Euclid more than two-thousand years ago to some of the more recent developments in mathematical physics. Published 30 years ago, MATHEMATICS: QUEEN AND SERVANT OF SCIENCE remains a valuable tool for teachers, students, and all those enthusiastic about mathematics. The book includes 20 chapters on selected areas in pure and applied mathematics, including discussions on topics such as geometry, algebra, logic, probability, infinity, transformations, rings, groups, topology, and continuity. Also covered are subjects such as: Mersenne primes, quantum mechanics, and three-valued logic. The book is not a history of mathematics, but rather an enticing sampling of selected topics in mathematics and science.

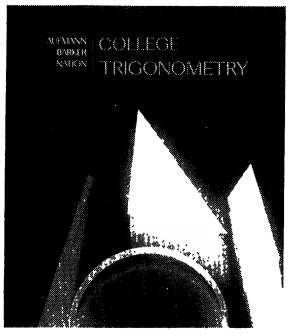
The book contains fascinating information on the personalities and philosophies of the great mathematicians, plus perceptive resumes of their specific contributions to the moving force of this great science.

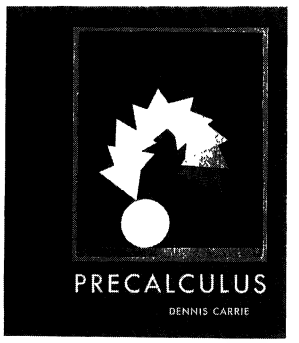


Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036









Factors in Success

College Algebra and Trigonometry
About 712 pages • hardcover

College Algebra
About 448 pages • hardcover

College Trigonometry
About 320 pages • hardcover

By **Richard N. Aufmann, Vernon C. Barker,** and **Richard Nation**
All of Palomar College

Study Guide • Instructor's Manual Student Disk • Computer Tutor™ • Test Booklet • Solutions Manual • Computerized Test Generator • Just published

The new Aufmann, Barker, Nation series helps students understand the interaction of theory and practice in mathematics. After each example, students work selected problems for a thorough grasp of key concepts. Applications from disciplines such as biology, business, economics, physics, archaeology, and psychology give meaning to problems and exercises.

Precalculus: An Introduction
Dennis Carrie, Golden West College

About 720 pages • hardcover • Instructor's Manual • Solutions Manual • Computerized Test Generator • Printed Test Bank • Just published

No other text prepares students for calculus as well as Carrie does. Elementary functions and graphing are covered early and reinforced throughout, developing essential skills for subsequent courses. Abundant word problems are based on real-life situations and range from routine to more conceptual manipulations.

For adoption consideration, request examination packages from your Houghton Mifflin regional office.



Houghton Mifflin

13400 Midway Rd., Dallas, TX 75244-5165
1900 S. Batavia Ave., Geneva, IL 60134
925 E. Meadow Dr., Palo Alto, CA 94303
101 Campus Dr., Princeton, NJ 08540

GRADUATE FELLOWSHIPS AT $12K PER ANNUM

Mathematical Sciences, University of Arizona

Up to 15 fellowships for outstanding new graduate students in the mathematical sciences will be available in 1990-91. Fellowship applicants should be seeking the Ph.D. and planning careers in teaching and/or fundamental research. Anticipated stipends are $12,000 for 12 months with both in-state and non-resident tuition waived.

While a few fellowships may be offered to foreign students, most are restricted to citizens, permanent residents, or individuals who have established intent to become citizens or permanent residents of the United States. Applications from U.S. women and students belonging to U.S. minority groups are particularly invited. Currently one-third of the U.S. graduate students in pure and applied mathematics at Arizona are women.

The University of Arizona has excellent programs in traditional pure and applied mathematics, and is a leading institution in interdisciplinary applied mathematics. This presents a wealth of opportunities for graduate study encompassing such areas as dynamical systems, number theory, computational science, computational group theory, nonlinear partial differential equations, mathematical physics, probability and problems in theoretical engineering and interdisciplinary applied physics. In addition, outstanding computational facilities for graduate study and research are available to the over 165 graduate students in the mathematical sciences at the University of Arizona.

Fellowship applicants of superior quality will be among the students invited to the Fourth Annual Workshop for Advanced Undergraduates on Current Ideas in the Mathematical Sciences, March 3-6, 1990. Limited support is available for attendees. The workshop is designed to communicate topics in current active research in three areas: (i) Geometry in Analysis and Number Theory, (ii) Numerical Methods for Multi-Scale Problems, and (iii) Mathematical Applications in the Life Sciences.

For information and application materials write: W. M. Greenlee or T. W. Secomb, Department of Mathematics/Program in Applied Mathematics, University of Arizona, Tucson, AZ 85721 (602) 621-2068.

The University of Arizona is an EOE/AA Employer

Arizona Center for the Mathematical Sciences
announces the
Fourth Annual Workshop for Advanced Undergraduates

CURRENT IDEAS IN THE MATHEMATICAL SCIENCES
March 3-6, 1990

The main goal of these workshops is to communicate to potential graduate students the level of excitement and activity in mathematical sciences today. The program will focus on topics in three areas:

1. Geometry in Analysis and Number Theory

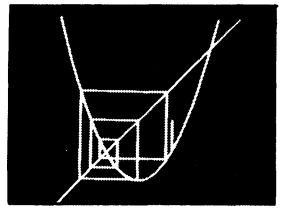
2. Numerical Methods for Multi-Scale Problems

3. Mathematical Applications in the Life Sciences

Partial travel and subsistence support is available. The deadline for application material is February 2, 1990. Direct all correspondence to:

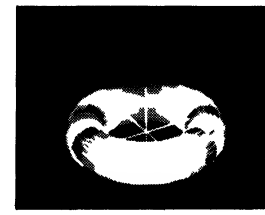
W. M. Greenlee or T. W. Secomb
Department of Mathematics, Program in Applied Mathematics
University of Arizona, Tucson, Arizona 85721
(602) 621-4664

EQUATIONS SYMBOLICALLY OR NUMERICALLY • HIDDEN-LINE SURFACE PLOTTING • 2-DIMENSIONAL DISPLAY OF ALGEBRAIC EXPRESSIONS • USER-DEFINED WINDOWS • EASY, MENU-DRIVEN INTERFACE • FULL-SCREEN TEXT EDITOR • DIFFERENTIAL EQUATIONS IN 2, 3, OR 4 DIMENSIONS • SUM SERIES • ALGEBRAIC SIMPLIFICATION RULES • SUM SERIES EXACT RATIONAL ARITHMETIC AND ALGEBRA • DECIMAL ARITHMETIC • SOLVE SYSTEMS OF



cal 2.0

discover it!



Are you wondering how you can use a computer algebra and graphics system in your mathematics classroom or laboratory?
We have some ideas that may surprise you.
cal 2.0, the computer algebra and graphics system for learning, offers you

- **Scripting:** Teachers create scripts - interactive lessons for your students
 A lecture, a laboratory, or a textbook on disk!
 Illustrate your points with high-resolution color pictures
 Include, as needed, mathematical objects or programs to support student exploration
- **Interactive Learning:** Students explore your scripts, experimenting with the ideas that you have introduced . . . *cal 2.0* is menu-driven, easy to learn
- **Slide Shows:** For group presentations, lectures, seminars
 Design your own graphics and text slides and save them as shows on disk

Don't take our word for it - see for yourself!
Order now, or ask for 30 day no obligation preview

cal 2.0
is field tested in calculus laboratory classrooms
is available for IBM AT or PS/2 and compatibles
requires 640K RAM and EGA or VGA graphics
specify EGA or VGA specify 3.5" or 5.25" disks



Bluejay Lispware
P.O. Box 1904
Gambier, Ohio 43022
(614) 427-3894
Price: $225.00
VISA

Take the problem out of problem solving

Beginning Algebra

Robert G. Marcucci
San Francisco State University
Harold L. Schoen, The University of Iowa

About 512 pages • hardcover • Student Solutions Manual • Instructor's Manual Answer Key • Computerized Test Generator • Printed Test Bank • Just published

Marcucci/Schoen introduces and reinforces mathematical concepts with a variety of real-world applications that effectively develop problem-solving capabilities. Case studies, boxed inserts, and numerous examples support text exposition, helping students solve the vast assortment of word problems.

Intermediate Algebra

Norman Siever
Los Angeles Valley College

About 672 pages • hardcover • Instructor's Manual • Student Solutions Manual Answer Key • Computerized Test Generator • Printed Test Bank • Just published

Intermediate Algebra reflects Siever's understanding of what skills students need to know and how to teach them. Using a friendly tone and carefully sequenced organization, Siever reinforces concepts with introductory outlines, boxed inserts, case studies, and clear examples and summaries. Siever creates a real-life context for learning by emphasizing problem-solving skills.

For adoption consideration, request examination packages from your Houghton Mifflin regional office.



Houghton Mifflin

13400 Midway Rd., Dallas, TX 75244-5165 1900 S. Batavia Ave., Geneva, IL 60134
925 E. Meadow Dr., Palo Alto, CA 94303 101 Campus Dr., Princeton, NJ 08540

GEOMETRY & TOPOLOGY

FIRST CONCEPTS OF TOPOLOGY: The Geometry of Mappings of Segments, Curves, Circles, and Disks

W.G. Chinn and N.E. Steenrod

This clear and winning little book, for readers willing to come to genuine grips with the idea of a mathematical proof, presents topology . . . as mathematicians see it. The parlor tricks are gracefully alluded to here and there, but they are distinctly for after hours. The center of interest is the stuff itself: powerful notions of set theory are . . . exploited to define open sets and their coverings, and from them [to prove] the key theorems. .. One cannot any longer doubt that a single stroke of a knife exists that divides any irregular ham sandwich so that ham and both bread slices can be shared with perfect fairness by two consumers.

Philip Morrison in *Scientific American*

Contains over one hundred and fifty problems and solutions.

160 pp., 1966, ISBN 0-88385-618-2

List: $10.00 MAA Member: $8.00

Catalog Number NML-18

FROM PYTHAGORAS TO EINSTEIN

K.O. Friedrichs

Starting with area and dissection proofs of the Pythagorean theorem, Friedrichs proceeds gently, step by step, to cover vectors, coordinates, elastic and inelastic impacts, and relativistic space-time–ending with a derivation of a contemporary formula rivaling the Pythagorean theorem in fame, $E = mc^2$.

88 pp., 1965, ISBN 0-88385-616-6

List: $9.50 MAA Member: $7.60

Catalog Number NML-16

GEOMETRY REVISITED

H.S.M. Coxeter and S.L. Greitzer

After seeing a proof that a segment has a unique midpoint, students may wonder whether there are any interesting theorems in geometry. If either they or their teachers get a hold of this book, they will learn that there are many beautiful and nontrivial theorems in geometry. Among those found here are the theorems of Ceva, Menelaus, Pappus, Desargues, Pascal, and Brianchon. A nice proof is given of Morley's remarkable theorem on angle trisectors. The transformational point of view is emphasized: reflections, rotations, translations, similarities, inversions, and affine and projective transformations. Many fascinating properties of circles, triangles, quadrilaterals, and conics are developed.

193 pp., 1967, ISBN 0-88385-619-0

List: $11.75 MAA Member: $9.50

Catalog Number NML-19

The very lucid presentation takes the reader from the elementary problems of plane Euclidean geometry to the fundamental concepts of non-Euclidean geometry . . . The book is rich in remarkable facts and thereby is very effective in promoting the significance and the value of geometry in mathematical teaching, a promotion which is very necessary . . . The always original developments use very simple tools . . . and soon proceed to higher configurations.

K. Strubecker in *Mathematical Reviews*

Order from:



The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036
(202) 387-5200

Writing Mathematics Well, by Leonard Gillman
64 pp., 1987, ISBN-0-88385-443-0
Catalog Number - WMW List: $6.50 MAA Member: $5.00

Good writing conveys more than the author originally had in mind, while poor writing conveys less. Well-written papers are more quickly accepted and put into print and more widely read and appreciated than poorly written ones—and for notes, monographs, and books the quality of writing is more important than it is for papers.

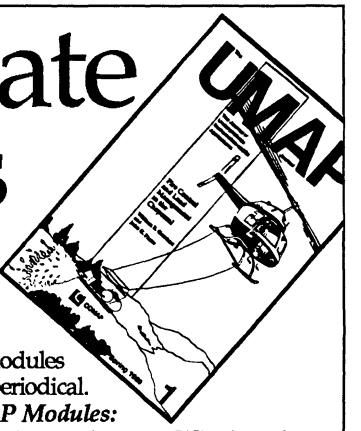
In **Writing Mathematics Well,** Leonard Gillman tells his readers how to develop a clear and effective style. All aspects of mathematical writing are covered, from general organization and choice of title, to the presentation of results, to fine points on using words and symbols, to revision, and finally, to the mechanics of putting your manuscript into print. No book can by itself make you a better writer, but this one will alert you to the opportunities for better and more forceful writing. It does this both by precept and by example.

A book to be read for its sharpness and wit as well as for enlightenment, **Writing Mathematics Well** should be on the shelf of anyone who writes or intends to write mathematics. It will amuse and delight the already careful writer and it will help reform and refine the sensibilities of those who may be somewhat careless about their writing.



Order from
The Mathematical Association of America
1529 Eighteenth St. NW
Washington, D.C. 20036

Undergraduate Applications From α to Ω



The UMAP Journal, a quarterly, blends up-to-date teaching modules with commentaries and articles to create a boldly different periodical. The *Journal* is complemented by the Annual Collection: *UMAP Modules: Tools for Teaching,* which contains new modules developed during the year. Whether it's an elegant explanation of price-elasticity in illegal drugs or the strategic implications of hiring a new secretary, we're constructing mathematical models to tackle the problems of a new age.

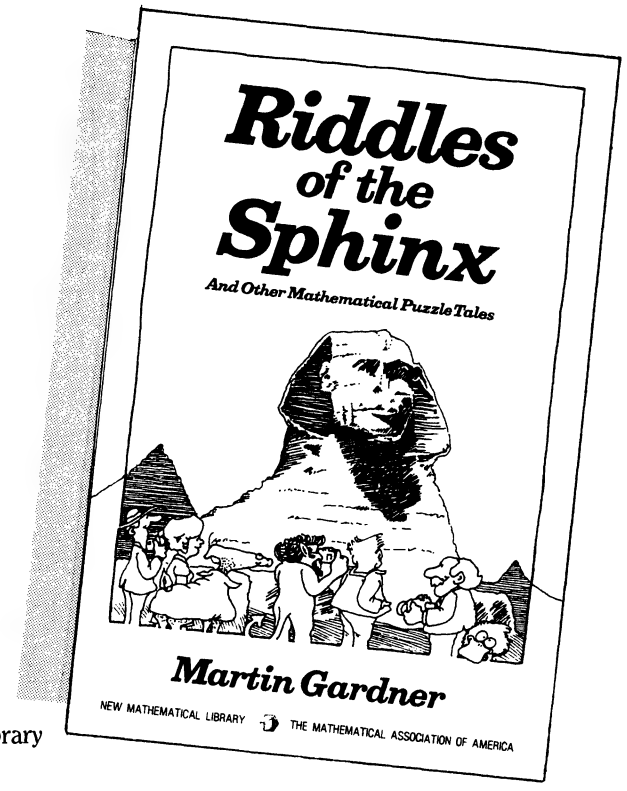
The UMAP Journal
Tackling The Problems Of A New Age

	Journal only	*Journal and UMAP Modules: Tools for Teaching*
Individual	$37.00/34.00* ($49.00/46.00* foreign)	$51.00/48.00* ($64.00/61.00* foreign)
Library Rate		$98.00 ($110.00 foreign)
Institutional (two copies per issue)		$148.00 ($160.00 foreign)

* Indicates discounted rate available to members of COMAP, AMATYC, MAA, NCTM, SIAM, ASA, and TIMS.

Riddles of the Sphinx

and other mathematical puzzle tales.



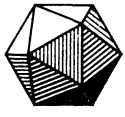
by Martin Gardner
Volume 32 in the New Mathematical Library
184 pp., Paper. ISBN-0-88285-632-8
List: $14.50 MAA Member: $12.50

Martin Gardner has charmed readers for over fifty years with his delightful books and articles on science. He is best known for the popular column, *"Mathematical Games,"* which appeared in *Scientific American* for twenty-five years. Generations of scientists and mathematicians have been inspired by his writing and the MAA is proud to include his name in its list of authors.

This book was drawn from Gardner's column in *Isaac Asimov's Science Fiction Magazine*. The riddles presented here incorporate the responses of his initial readers, along with additions suggested by the editors of the New Mathematical Library. Each chapter (riddle) poses a problem answered in the First Answers section. The solution in turn raises another problem that is solved in the Second Answers section. This may suggest a third question and in several instances there is a fourth. Gardner draws us from questions to answers always presenting us with new riddles—some as yet unanswered. There are 125 different pieces altogether.

Solving these riddles is not simply a matter of logic and calculation, although these play a role. Luck and inspiration are factors as well, so beginners and experts alike may profitably exercise their wits on Gardner's problems, whose subjects range from geometry to word play to questions relating to physics and geology.

We guarantee that you will solve some of the riddles, be stumped by others, and be amused by almost all of the stories and settings that Gardner has devised to raise these questions.



Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036



Publications by the
Mathematical Association of America
1529 Eighteenth St., NW
Washington, DC 20036
202 387-5200

New Mathematical Library #30 . . .

The Role of Mathematics in Science
by Leon Bowden and M.M. Schiffer
220 pp. Paperbound
List: $16.00 MAA Member: $13.00

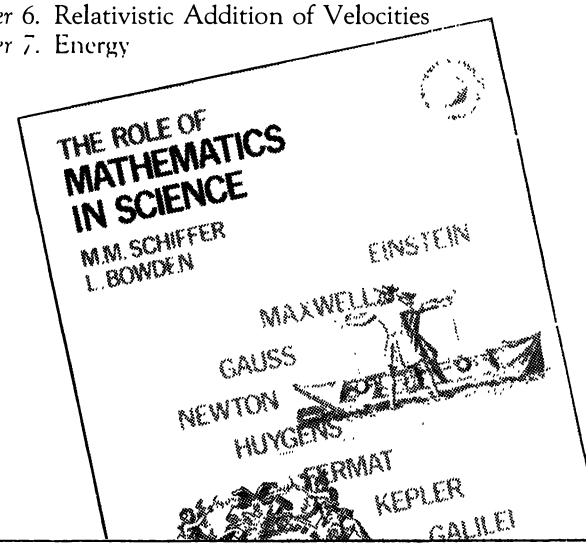
This book is based on a series of lectures given over a period of years to high school mathematics teachers. The lectures focused on topics in mechanics, questions of population growth, probability and other uses of exponential functions, optics and application of matrices to relativity theory.

The aim of the book is to illustrate the power and elegance of mathematical reasoning in science with some examples ranging from the work of Archimedes to that of Einstein. The book starts with problems of the lever, the mirror and the growth of populations and ends up with problems of space travel and atomic energy.

This excellent book will appeal to the advanced high school student, to the undergraduate and to their teachers. A sampling of the Table of Contents should inspire you to order your copy now.

Table of Contents

Chapter 1. The Beginnings of Mechanics
Chapter 2. Growth Functions
Chapter 3. The Role of Mathematics in Optics
Chapter 4. Mathematics with Matrices - Transformations
Chapter 5. What is the Time? Einstein's Transformation Problem
Chapter 6. Relativistic Addition of Velocities
Chapter 7. Energy



MAA STUDIES IN MATHEMATICS

Studies in Numerical Analysis

MAA Studies in Mathematics #24

Gene H. Golub, Editor

415 pp. Hardbound.

List: $46.50 MAA Member: $34.50

This volume is a collection of papers describing the wide range of research activity in numerical analysis. The articles describe solutions to a variety of problems using many different kinds of computational tools. Some of the computations require nothing more than a hand held calculator: others require the most modern computer. While the papers do not cover all of the problems that arise in numerical analysis, they do offer an enticing and informative sample.

Numerical analysis has a long tradition within mathematics and science, beginning with the work of the early astronomers who needed numerical procedures to help them solve complex problems. The subject has grown and developed many branches, but it has not become compartmentalized. Solving problems using numerical techniques often requires an understanding of several of the branches. This fact is reflected in the papers in this collection.

Computational devices have expanded tremendously over the years, and the papers in this volume present the different techniques needed for and made possible by several of these computational devices.

Table of Contents

The Perfidious Polynomial, *James H. Wilkinson*

Newton's Method, *Jorge J. Moré and D. C. Sorensen*

Research Directions in Sparse Matrix Computations, *Iain S. Duff*

Questions of Numerical Conditions Related to Polynomials, *Walter Gautschi*

A Generalized Conjugate Gradient Method for the Numerical Solution of Elliptic Partial Differential Equations, *Paul Concus, Gene H. Golub and Dianne P. O'Leary*

Solving Differential Equations on a Hand Held Programmable Calculator. *J. Barkley Rosser*

Finite Difference Solution of Boundary Value Problems in Ordinary Differential Equations, *V. Pereyra*

Multigrid Methods for Partial Differential Equations, *Dennis C. Jespersen*

Fast Poisson Solvers, *Paul N. Swarztrauber*

Poisson's Equation in a Hypercube: Discrete Fourier Methods, Eigenfunction Expansions, Pade Approximation to Eigenvalues, *Peter Henrici*



Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 2 February 1990

Contents

(ISSN 0002-9890)

ARTICLES

A Visit to the Newtonian N-body Problem via Elementary Complex Variables DONALD G. SAARI 105

The Square Pyramid Puzzle W. S. ANGLIN 120

EDITOR'S CORNER

The Euclidean Algorithm Strikes Again STAN WAGON 125

LETTERS TO THE EDITOR 130

UNSOLVED PROBLEMS

Open Problems in Grid Labeling JOSEPH A. GALLIAN 133

NOTES

A Theorem on the Numerators of the Bernoulli Numbers KURT GIRSTMAIR 136

Another Proof of the Quadratic Reciprocity Theorem? RICHARD G. SWAN 138

Adding up to Powers EDWARD A. BENDER, FRED KOCHMAN, AND DOUGLAS B. WEST 139

THE TEACHING OF MATHEMATICS

A One-Sentence Proof That Every Prime $p \equiv 1 (\mathrm{mod}\, 4)$ Is a Sum of Two Squares D. ZAGIER 144

Inverse Functions and their Derivatives ERNST SNAPPER 144

Extending the Principal Axis Theorem to Fields Other Than $\mathbb{R}$ STEPHEN H. FRIEDBERG 147

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 150

Advanced Problems and Solutions 161

REVIEWS

Representations of *-Algebras, Locally Compact Groups, and Banach *-Algebraic Bundles By J. M. G. Fell and R. S. Doran MARC A. RIEFFEL 167

TELEGRAPHIC REVIEWS 171

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

A Visit to the Newtonian N-body Problem via Elementary Complex Variables

DONALD G. SAARI, *Northwestern University, Evanston, Il 60208*



DON SAARI received his undergraduate degree from Michigan Technological University and his Ph.D. in mathematics from Purdue University. His Ph.D. dissertation, written under the guidance of Harry Pollard, concerned the qualitative behavior of dynamics in the general Newtonian N-body problem. After a postdoctoral position in the Yale Astronomy Department, Saari moved to Northwestern University where he currently is Professor of Mathematics. His main research interests revolve around applications of dynamical systems to celestial mechanics and physics and to issues coming from the social sciences.

The study of how N celestial bodies move under gravitational forces is an old one. If one is willing to acknowledge the work of the ancient astrologers and shepherds—two groups that carefully plotted the positions of the stars and planets —then this subject area traces its origins to the earliest reaches of mankind. Indeed, had the title not been already preempted, one might suggest that the study of the N-body problem is "the world's oldest profession." If it isn't the oldest, then, most surely, it is "the second oldest."

How do the heavenly bodies move? With the help of elementary complex variables, certain selective orbits will be described where, as it turns out, even "simple" motion can be surprisingly complicated. Also, some of the history of the Newtonian N-body problem will be related with an emphasis on the myth that only the two body problem has been solved. For practical purposes this assertion is correct, but at the turn of the century the Finnish mathematician K. Sundman "solved" the three body problem in an accepted sense of that time. While explaining why Sundman's result isn't widely known, it will be indicated how collisions, both the real type where two or more particles hit each other and the complex type where imaginary collisions occur at complex values of time, affect the behavior of the system.

To start, consider the question that, at some time, probably bothered many of us. Namely, why did the early astronomers have so much trouble predicting the motion of the planets? Armed with a 20th-century education, we know that a reasonable approximation for the motion of a planet is uniform circular motion about the Sun. What is so difficult about describing motion as predictable as this? This question is closely related to the fable where Galileo silently disavowed his forced recanting of a Sun-centered solar system. Probably generations of irreverent school children silently wondered, "Who cares? What difference does it make if the Sun or if the Earth is the center?" It does matter. This simple change of variables introduces a significantly different perspective of the solar system, it explains the difficulties faced by the astronomers of antiquity, and it underscores the critical importance of the Copernican revolution.

To understand why an Earth-centered prejudice creates problems, consider a simplified version of the Sun-Earth-Mars system. Mars is approximately $3/2$ times as far from the Sun as the Earth (1.524 times) and it takes approximately 2 years

(687 Earth days) to complete one revolution of the Sun. To eliminate fractions, replace the traditional astronomical unit with half-astronomical units so that, in the new system, the Earth is distance 2 from the Sun. In this simplified model, the Earth's position is $z_E(t) = 2e^{2\pi it}$ where t is measured in "Earth years," while that of Mars is $z_M(t) = 3e^{\pi it}$. The position of Mars as observed from Earth is

$$z(t) = z_M(t) - z_E(t) = 3e^{\pi it} - 2e^{2\pi it}. \tag{1.1}$$

A convenient way to describe this orbit is with the translation

$$z(t) - 2 = 3e^{\pi it} - 2e^{2\pi it} - 2 = e^{\pi it}(3 - 2e^{\pi it} - 2e^{-\pi it}) = (3 - 4\cos \pi t)e^{\pi it}$$

The graph of this equation—the polar coordinate representation of a limaçon with a loop—depicts the orbit of Mars relative to Earth.[1] (See FIGURE 1.)

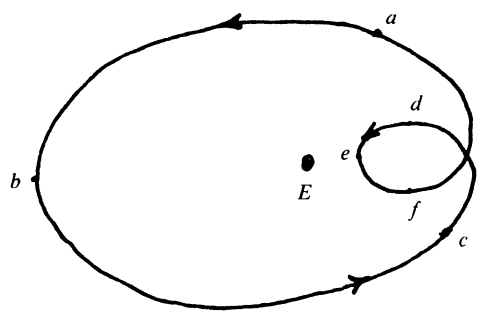


FIG. 1.

It is clear from FIGURE 1 why the pre-Copernican, Earth-centered prejudice made it so difficult to predict the motion of the planets and to develop a "Newtonian Theory." While the orbit from a to b to c is not overly complex, it becomes quite complicated once Mars passes through point c. Mars continues in a counterclockwise direction until point d where it appears to *stop* and then change to a clockwise motion until point f. At f, Mars again reverses direction to return to a counterclockwise motion. Imagine what complications this motion presented to astronomers trying to predict positions. Next, imagine a theorist trying to determine the governing laws of motion. While Mars goes from a to b to c to d, a theorist might persuasively argue for a law of attraction. But, how does one explain the orbit from e to f to a? How does one justify the sudden law of repulsion? What is there about the Earth that drives Mars away?

Actually, the motion of Mars in FIGURE 1 is simpler than that of the other planets. Similar elementary complex variable arguments prove that the orbits of Saturn, of Jupiter, and of the other planets that take much longer to circle the Sun, admit several loops similar to the one in FIGURE 2. This can be proved as above, but

[1] The trigonometric version of this argument involves nothing more difficult than the double angle formulas, so it can be used to motivate several topics in calculus and precalculus classes. For instance, I find that this description of the orbit of Mars serves as a more persuasive illustration of the relevance of limaçons and cardioids than many of the standard examples used in calculus.

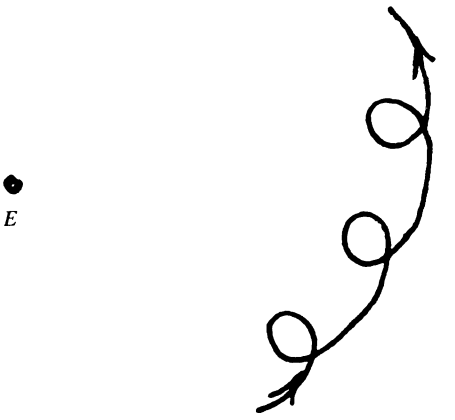


FIG. 2.

now the limaçon is relative to a rotating circle. (Instead of finding the orbit relative to a fixed translation, find it relative to a rotating one chosen so that the exponential terms collapse to a cosine term.) As an alternative way to demonstrate this, express the relative orbit, $z(t) = ae^{\alpha\pi it} - 2e^{2\pi it}$, $a > 2$, in the standard complex form of $z(t) = r(t)e^{i\theta(t)}$. The direction of the motion, whether it is clockwise or counterclockwise, is determined by the sign of θ'. This is the imaginary part of $(\ln z(t))' = z'/z = r'/r + i\theta'$. But $z'/z = \pi i(a\alpha - 4e^{(2-\alpha)\pi it})/(a - 2e^{(2-\alpha)\pi it})$, so after an elementary computation, it follows that the sign of θ' changes if $a\alpha < 4$. This inequality holds because Kepler's third law asserts that $a^3\alpha^2 = k$ where k is a constant; thus $a\alpha = (k/a)^{1/2}$ is a decreasing function of a.

Pictures of these looping orbits depicting the motion of the distant planets relative to the earth can be found in several books on the history of astronomy. Even more impressive are the photos of the planets taken over a several year period that are superimposed on one plate. When one compares this kind of dynamical behavior with the uniform motion obtained through a Sun centered system, it becomes clear why the Copernican change of variables has had such a profound, simplifying impact on science and astronomy.

The Copernican representation significantly simplifies the description of the motion of the planets, but is it a correct theory? This is the essence of Cardinal Barberini's query when he contested Galileo (at least in Brecht's play *Life of Galileo*), "*Are you sure... you astronomers aren't just trying to make astronomy a little easier for yourselves?... You like to think in circles or ellipses and in uniform velocities, in simple motions commensurate with your minds. But what if God had been pleased to make His stars move like this?*" where Barberini moves his finger through the air in a complicated course that presumably resembles FIGURE 2. Responding with his version of Occam's razor, Galileo argued for Copernicus's theory by asserting, "*if God had created the world like this* [He retraces Barberini's course.] *He would have constructed our minds like this too* [He repeats the same course.] *to enable them to recognize these courses as the simplest. I believe in reason.*"

Elementary complex variables disclose these somewhat surprising, relative orbits for the outer planets by capturing the effects of the rotating coordinate system defined by the motion of the Earth. There are other natural rotating systems in

astronomy, so one might correctly suspect that a related approach would exhibit other surprising astronomical behavior. To demonstrate this, consider the planet Mercury. It takes Mercury about 90 Earth days (87.967) to circle the Sun. Based on observations, it was believed until the 1960s that Mercury took about the same length of time to rotate on its axis. If this were true, then, as asserted in most of the older texts on astronomy, the same face of Mercury always would face the Sun—a solar day (or solar night) on Mercury would last forever. However, with radar observations, we now know that Mercury takes about 60 Earth days to rotate on its axis; this shortens a Mercury solar day to about 176 Earth days, or two Mercury years. Moreover, as shown next, when a more accurate representation for the orbit of Mercury is used, an observer on Mercury would find the apparent orbit of the Sun to be quite peculiar.

To obtain a sharper approximation for the orbit of Mercury, or any other planet about the Sun, treat the orbit as an ellipse with eccentricity ε; e.g., for the Earth, $\varepsilon = 0.0167$, while for Mercury, $\varepsilon = 0.2056$. The position of the planet on the ellipse is given by

$$r(\theta) = a/(1 - \varepsilon\cos(\theta)) \approx a(1 + \varepsilon\cos(\theta)), \tag{1.2}$$

where a is a positive constant and $\theta(t)$ is the angle of the planet determined by a reference line. Equation 1.2 has the complex representation $z(\theta) \approx a(1 + \varepsilon\cos(\theta))e^{i\theta} = a(1 + (\varepsilon/2)e^{i\theta} + (\varepsilon/2)e^{-i\theta})e^{i\theta}$, or

$$z(\theta) \approx a\varepsilon/2 + ae^{i\theta} + (a\varepsilon/2)e^{2i\theta}. \tag{1.3}$$

The rotation of Mercury is given by $e^{12\pi i}$. It follows that the apparent position of the Sun on Mercury is

$$Z(t) = -z(\theta(t))e^{-12\pi it}, \tag{1.4}$$

so

$$\text{Arg}(Z(t)) = \text{Arg}(z(\theta(t)) - 12\pi t + \pi = \theta(t) - 12\pi + \pi. \tag{1.5}$$

The apparent motion of the Sun changes directions whenever $\text{Arg}(Z(t))'$ changes sign. This happens, and the reason is based on Kepler's second law, which asserts that $r^2\theta'$ is constant valued. Thus, if ε is sufficiently large, as it is for Mercury, $\theta(t)$ cannot be represented by uniform motion. In particular, when r is at perihelion (i.e., $r(t) = |z(t)|$ is at a minimum), θ' is at its maximum value. Indeed, using Kepler's second law and eq. 1.2, we have that $\theta' = n/(1 - \varepsilon\cos(\theta))^2$ where n is the mean motion; e.g., for Mercury, $n = 8\pi$. Consequently, at perihelion, $\theta' = 8\pi/(0.7944)^2 > 12\pi$. This same argument shows that $\text{Arg}(Z(t))' \geq 0$ if $|\theta| \leq 26.8°$; otherwise it is negative. From this, a schematic representation of the apparent motion of the Sun, given in FIGURE 3, is easy to determine. Each of the two loops corresponds to Mercury's "yearly" passage through perihelion. (A similar argument shows that the apparent motion of the Sun would have no reversal of direction had the orbit of Mercury been circular enough so that $\varepsilon < 0.1835$.)

The change in the sign of $\text{Arg}(Z(t))$ creates an interesting phenomenon. There are locations on Mercury where after the Sun rises in the east on a Mercury morning, it reaches only a certain point in the sky before it *stops* to retrace its motion and to set in the *east*. Then, the Sun rises a second time this "morning," but this time it progresses normally throughout the long day. That's not all; after it sets

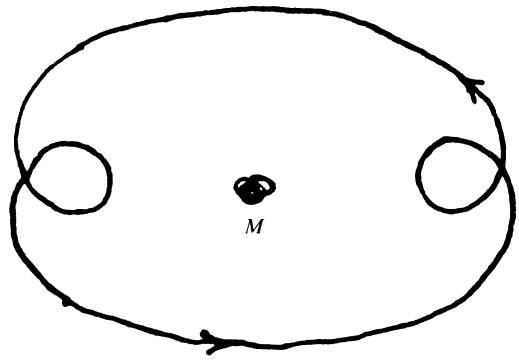


FIG. 3.

in the *west*, this changeable Sun reverses behavior once more to rise again in the west for a brief time before it finally sets for the long Mercury night.

2. Epicycles. An important planetary theory was developed by Ptolemy. One can appreciate the genius of Ptolemy by examining the orbits depicted in FIGURES 1 and 2. Recall the problem facing him: a scientific theory of planetary motion needed to satisfy the governing prejudices of the time. For Ptolemy, this meant a theory needed to accommodate the Earth as the center of the solar system, and we've just seen the complexities associated with this assumption. The next obstacle, left over from Aristotle, was that the circle is the most perfect figure in geometry. Obviously, heavenly bodies are "incorruptible," so their motions should be described by uniform circular motion. But how does one build a predictive, planetary theory incorporating these assumptions?

In his *Almagest*, written around A.D. 130, Ptolemy invented the ingenious approach of epicycles. This is where the position of the particles satisfies Aristotle's constraint of uniform circular motion. The clever idea is that the point indicated by the motion of the first circle, the deferent, does *not* represent the location of the planet. It locates the center of a *second* circle spinning with uniform motion. The location of the planet is given by the moving point on the second circle—the epicycle. (See FIGURE 4.) Today this approach may seem to be hopelessly naive, but remember that variations of this theory dominated astronomy for more than a millennium—an incredibly long time for any scientific theory. Even Newton's

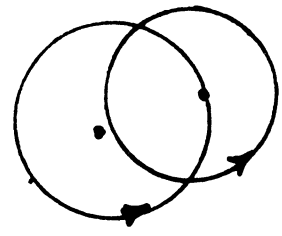


FIG. 4.

equations didn't enjoy such a long reign before being challenged by Einstein's relativity.

The long success of Ptolemy's approach can be understood by using complex variables. Let a_j be the radius of the jth circle where the period of the uniform motion is $1/2b(j)$, $j = 1, 2$. This means that the motion of a planet, as described through epicycles, is given by

$$z(t) = a_1 e^{b(1)\pi it} + a_2 e^{b(2)\pi it}. \tag{2.1}$$

From this equation, the source of the success of the epicycles becomes obvious. Namely, with the appropriate choices of a_j and $b(j)$, $j = 1, 2$, eq. 2.1 is the same as eq. 1.1. In other words, the epicycle approach can be viewed as recapturing the change of variables used to convert the Sun centered motion of a planet to an Earth centered one. With the appropriate choice of constants, this epicycle expression is the correct one for any planet.

As time marched on, the accuracy of Ptolemy's theory didn't always satisfy the increasing demands of astrology and astronomy. In part, this is because the actual orbits are on perturbed ellipses rather than on circles about the Sun. (Some of Ptolemy's orbits already incorporated elliptic behavior.) To achieve more accurate results, modifications of the same idea could be used. The approach is simple: instead of treating the location of the moving point on the second circle as the location of the planet, treat it as the center of a third (or fourth, or fifth, or ...) circle moving in uniform circular motion. The complex variable representation of these approximations is

$$z(t) = \sum_j a_j e^{b(j)\pi it} \tag{2.2}$$

where $j = 1, \ldots, k$, and k is the number of circles being used. As already illustrated by eq. 1.3, such an approach can achieve a higher degree of accuracy. Also, imagine how this never ending problem of determining the next values of a_k and $b(k)$ could serve as the source of an infinite number of Ph.D. dissertations.

Epicycles were abandoned long ago. We now know, after extensive development of sophisticated mathematical theories applied to Newton's equations, that there are situations where the orbits of planets are either quasi-periodic or almost periodic. What is quasi-periodic motion? It is motion represented by eq. 2.2, the epicycles, while almost periodic motion is the limit of this summation as $k \to \infty$.

3. Collisions and spinors. Today, Newton's equations are used to study the N-body problem. One approach to solving these equations might be to find the series solution. As we know from complex variables, the radius of convergence for this series is determined by the distance to the nearest singularity. Such a singularity appears to be a collision. (For $N \geqslant 4$, the situation is more complicated; see [**13**].) Thus, as a first step toward finding series solutions of Newton's equations, we need to understand the properties of collisions. For instance, one might ask whether it is possible to mathematically continue a solution through a collision? If so, this could extend the radius of convergence of a series.

To develop the behavior of collisions, start with the one body, or central force problem. If $\mathbf{r}(t)$ represents the vector position of the particle, then, with the

appropriate units of mass, time, and distance, the equations of motion are

$$\mathbf{r}(t)'' = -\mathbf{r}/r^3, \tag{3.1}$$

where $r = |\mathbf{r}|$. The solutions, given by eq. 1.2, are ellipses, parabolas, or hyperbolas. But, what happens near or at a collision? Can we define the motion through $r = 0$? This question was posed and answered by Sundman [**16**] in 1913. An improved, simplified theory was developed by the Italian geometer Levi-Civita [**4**] in 1920.

To see the ideas, start with a family of orbits that have a collision in the limit, i.e., a family represented by eq. 1.2 where ε tends to unity. Such a family is depicted in FIGURE 5. The key point is that when a particle has a "close approach," it rapidly spins around the central body. Therefore, if the dynamics can be mathematically extended beyond a collision, the colliding particle must make a 2π angular change —we should expect the particle to hit and then rebound from the central force. To remove this collision singularity from the equations of motion, this abrupt change needs to be removed by straightening out the orbit.

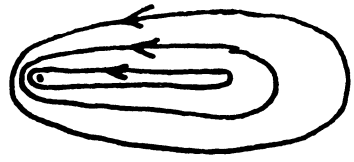


FIG. 5.

One way to see how to do this is to restrict the motion to the plane. Reexpress the position vector $\mathbf{r} = (x, y)$ as a complex variable $z = x + iy$, so the equations of motion are $z'' = -z/r^3$, where $r = |z|$. The collision occurs at $r = 0$, so this is where a change of variables is needed to convert the abrupt, 2π polar angle change into a form where the motion is on a straight line. To do this, the change of dependent variables must take half of the angular change at this point. With the complex variable representation of the coordinates, the appropriate change of variables is obvious; use the square root. The coordinate change is $w = u_1 + iu_2 = z^{1/2}$, or

$$w^2 = z. \tag{3.2}$$

The equations of motion for w, when accompanied with the change of independent variables $ds = dt/r(t)$ introduced by Sundman, not only are well defined at $w = 0$, a collision, but they assume the particularly simple form

$$\mathbf{u}'' + a\mathbf{u} = \mathbf{0} \tag{3.3}$$

where a is some positive constant and $\mathbf{u} = (u_1, u_2)$ is the vector representation for w. In other words, *this transformation converts the nonlinear Newtonian equations into the linear equations for a harmonic oscillator.*

To see how eq. 3.3 arises, note that the change of the independent variable defines the operator $d^2/dt^2 = [rd^2/ds^2 - r'd/ds]/r^3$, where $(')$ is differentiation with respect to s. An important contribution to the derivation of eq. 3.3 is the r^3

term in the denominator of the operator; it cancels the r^3 term in the Newtonian force. Thus, the equations of motion and the energy integral ($\mathbf{v}^2 = 2(r^{-1} + h)$) are, respectively,

$$rz'' - r'z' + z = 0, \qquad |z'|^2 = 2(r + hr^2) \tag{3.4}$$

where h is the constant of energy. When the new dependent variable is included, a cancellation of terms occurs if r, r' are replaced with their representation in terms of w and its conjugate. The new equations of motion and energy integral are

$$2w'' - \frac{w}{r}(2|w'|^2 - 1) = 0, \qquad 2|w'|^2 = 1 + hr. \tag{3.5}$$

Making the obvious substitution with the energy integral, the equations of motion become $w'' - (h/2)w = 0$. Thus, both this substitution and the coefficient $a = -h/2$ in eq. 3.3 reflect the fact that the transformed equations are on a fixed energy surface. Because the change of the dependent variables is not 1-1, two $\mathbf{u}$ values correspond to a single $\mathbf{r}$ value. The exception is $\mathbf{u} = \mathbf{0}$ which corresponds to $\mathbf{r} = \mathbf{0}$.

In addition to being able to analyze the behavior of orbits near a binary collision, another advantage of eq. 3.3 is that they are linear equations with purely complex eigenvalues. From this it follows that all solutions are stable; a small perturbation has only a small effect on the solution. As such, numerical solutions of these equations retain the properties of the actual solution. This isn't true for eq. 3.1. Here, a small numerical error can force the numerical solution onto a different energy surface, which, in turn, alters the frequency of the motion. Then, as is true for two pendulums with close, but different frequencies, the true and the computed solutions eventually will be at opposite ends of the orbits.

Several important dynamical systems, such as the Earth satellite problem, are perturbed forms of eq. 3.1. (The perturbations for the equations of the Earth satellite problem reflect the Earth's "middle age bulge" around the equator.) Many of these systems cannot be solved analytically; instead they are numerically integrated. Consequently, it is natural to question whether perturbed forms of eq. 3.1 also can be converted into a form that inherits some of the stability properties of the harmonic oscillator. However, "real" problems, such as the Earth satellite problem, are in a three-dimensional space; they can't be restricted to the "Flatland" setting of a fixed two-dimensional plane. Therefore, to carry out the program of converting the satellite problem into a perturbed form of the three-dimensional harmonic oscillator, one first must be able to convert the three-dimensional eq. 3.1 into a three-dimensional harmonic oscillator. It appears from comments in the literature (e.g., see [**15**, p. 23]) that Levi-Civita unsuccessfully tried to find such a three-dimensional extension of his two-dimensional solution.

As the world moved into the Space Age, this issue of finding a harmonic oscillator form for the three-dimensional, two-body problem added practical importance to its theoretical interest. This question was raised by E. Stiefel at the 1964 Oberwolfach conference he organized. Attending this conference was P. Kustaanheimo, a Finnish astronomer, who had been using spinors to analyze properties of his theory of relativity and other problems from physics. Spinors are a natural generalization of complex variables, and, sure enough, by mimicking Levi-Civita's approach with spinors, Kustaanheimo solved the problem during this conference. More specifically, let $w = u_1 + iu_2 + ju_3 + ku_4$, $z = x_1 + ix_2 + jx_3 +$

kx_4, and $\mathbf{u} = (u_1, u_2, u_3, u_4)$. With the change of dependent variables, $w^2 = z$, along with Sundman's change of the independent variable, Kustaanheimo [2] converted the system into the equations

$$\mathbf{u}'' + a\mathbf{u} = 0. \tag{3.6}$$

Of course, the true problem is in R^3, not in R^4. Fortunately, as indicated later, this change of variables maps each solution of eq. 3.6 to a fixed plane $x_4 = c$. Thus, x_4 is a dummy variable that can be set equal to zero, and solutions of eq. 3.6 are mapped to solutions of eq. 3.1 in the plane $x_4 = 0$. Moreover, because a point z in the plane $x_4 = 0$ determines the magnitude of the corresponding $\mathbf{u}$, it follows from the dimensional difference that each point in the plane $x_4 = 0$ corresponds to a circle in the $\mathbf{u}$ variables rather than the two points in the Levi-Civita transformation. The only exception is when $\mathbf{r} = \mathbf{u} = \mathbf{0}$, where this circle degenerates into a point.

Now that the spinor transformation solved this decades old problem, Kustaanheimo and Steifel [3] joined forces to use this approach to analyze perturbed problems of equation 3.1. Among other conclusions, they showed that this transformation does provide numerical advantage. (Extensions are in [15].) This transformation, which is quite widely used today (more so in Europe than in the U.S.), is known as the KS transformation.

What was the source of the difficulties in extending Levi-Civita's approach? Why can't we find a similar relation for R^3? One way to see this is to express Levi-Civita's change of dependent variables, $x = u_1^2 - u_2^2$, $y = 2u_1u_2$ in the differential form

$$\begin{pmatrix} dx \\ dy \end{pmatrix} = 2\begin{pmatrix} u_1 & -u_2 \\ u_2 & u_1 \end{pmatrix}\begin{pmatrix} du_1 \\ du_2 \end{pmatrix}. \tag{3.7}$$

The first column in this orthogonal matrix can be viewed as locating a point in space, say, on the unit circle $u_1^2 + u_2^2 = 1$, while the second column defines a tangent vector. Thus, the Levi-Civita transformation can be identified with a vector field of unit length on S^1—the unit circle in R^2. Presumably, the appropriate R^3 change of dependent variables leading to the harmonic oscillator would be given by a 3×3 orthogonal matrix where the first column locates a point on S^2, the unit ball in R^3, while the remaining two columns correspond to tangent vectors to the sphere. But the existence of any such matrix is frustrated by the "hairy billiard ball" effect—one can't comb a hairy billiard ball without getting a "cowlick" where at least one hair stands upright. In other terms, as we know from Poincaré and Brouwer, there doesn't exist a smooth, nonzero, tangent vector field on S^2. In our setting, this means there doesn't exist a 3×3 orthogonal matrix of the appropriate form, or, in turn, these topological reasons proscribe the existence of the desired change of variables.

On the other hand, going back to his Ph.D. dissertation and his earlier work in algebraic topology, E. Stiefel [15, see Chp 11] was aware of the fact that while it is impossible to comb a hairy S^2, one can comb a hairy S^3, the unit ball in R^4, in several ways. This unit ball admits a smooth frame where if (u_1, u_2, u_3, u_4) determines a point on S^3, then the remaining three mutually orthogonal, unit tangent vectors are $(-u_2, u_1, u_4, -u_3)$, $(-u_3, -u_4, u_1, u_2)$, and $(u_4, -u_3, u_2, -u_1)$. Once these four vectors are expressed as columns of a 4×4 matrix, much as in eq. 3.7, this defines the KS transformation of the dependent variables. Note that the last

row of this matrix is orthogonal to $\mathbf{u}$; this ensures that $x_4 = 0$. Indeed, the resulting mapping is one of the Hopf maps from S^3 to S^2.

Subsequent to the KS transformation, other representations of the two-body motion have been found to eliminate the difficulties of collisions. A geometric approach is given by J. Moser [7], where he showed that two-body motion in a d-dimensional physical space can be related to geodesic flow on S^d. Thus in the specific case $d = 3$, the problem in R^3 is related, in a different manner, to motion on S^3. In Moser's representation, the positions on the sphere correspond to the velocity of the particle through stereographic projections. Hence, the north pole of S^d represents the collision where the velocity becomes infinite. J. Milnor [6] has a nice exposition of this and related ideas. A second important development is due to R. McGehee [8]. Remember, Sundman and Levi-Civita found transformed equations that included the point $\mathbf{r} = \mathbf{0}$ as a regular point. These approaches don't extend for triple collisions. However, McGehee found a way to "paste on" the behavior of orbits at triple collisions when $r = 0$. He did this by expressing the equations of motion in a spherical coordinate framework, and by introducing an appropriate change of independent variable. The next step was to exploit the singularity in spherical coordinates that occurs when the radius is zero but all of the angles are admitted. In his representation, the "angles" correspond to the configurations formed by the particles while the radius of the sphere measured the distance between colliding particles. In this manner, the motion in the collinear three-body problem can be extended to a "collision manifold" corresponding to fictitious motion where the radius of the system is zero; i.e., this is the limiting motion when the distances between the particles approaches zero. R. Devaney [1] has a nice exposition of these ideas.

4. Sundman theory. At the beginning of this century, K. Sundman made several important contributions to our understanding of the N-body problem—contributions that have stood the test of time. Ironically, one of his major conclusions killed interest in a line of inquiry, so this particular result is not very well known. It should be; it is where Sundman "solved" the three-body problem according to accepted standards of the late 1800s and early 1900s. Indeed, in the late 1800s the King of Sweden and Norway established a prize for anyone who could find the solution of the N-body problem. The prize was awarded to Poincaré in 1889 even though he hadn't solved the original problem. (On the other hand, Poincaré's prizewinning paper contains a wealth of ideas that remain influential.) The originally stated problem finally was solved in 1913 by Sundman [16] when he found a converging series solution for the three-body problem. Unfortunately, his series converges so slowly that, essentially, it is useless for any practical purpose. Consequently, this nice result is not as well known as it should be. Still, Sundman's work remains a beautiful and important combination of several deep ideas that have a continuing influence on celestial mechanics. Sundman's results can be described via complex variables, so these variables will be used to outline what Sundman did and why his series converges so slowly. Part of this description of Sundman's contributions is based on Cauchy's theorem showing that the radius of convergence of a power series is determined by the location of the nearest singularity.

A serious complication hindering achievement of a convergent power series solution for the three-body problem are the singularities caused by binary collisions.

We now know that collisions of any kind are improbable [**11**]. But collisions exist, and when they do, they restrict the accompanying radius of convergence for any power series solution. One way to avoid these complications might be to restrict attention to those conditions with collision-free orbits. The fault with such an approach is that we don't know how to characterize these conditions; in general, we don't know whether an initial condition will, or will not lead to a collision-free trajectory.

As already indicated, Sundman cleverly avoided the problem of binary collisions by converting the equations to a new system where a binary collision is *not* a singularity—it becomes a regular point. Because of the central role played by this result, another complex variable argument will be outlined to suggest why we should expect this conclusion and why Sundman's change of independent variable is a "natural" one.

To start, consider colliding particles in the collinear, central force problem. The motion is on the positive half of the x-axis so the defining equations of motion are $x'' = -x^{-2}$. Because the right-hand side of this equation always is negative, the solution, $x(t)$, is concave down. This forces the system to have a collision either forwards or backwards in time, and if $x'(t) \leqslant 0$, then x' remains negative in the future. To determine how the collision occurs, multiply both sides of the equations of motion by x' and integrate to obtain the energy integral

$$(x')^2 = 2x^{-1} + 2h, \tag{4.1}$$

where h is a constant of integration.

If there is a collision at time t_0, then $x \to 0$ as $t \to t_0$. This changes the energy integral to $x(x')^2 = 2 + 2hx \sim 2$. In turn, this means that $x^{1/2}x' \sim -2^{1/2}$, or that $x^{3/2}(t) \sim A(t_0 - t)$. Consequently,

$$x(t) \sim A(t_0 - t)^{2/3} \qquad \text{as } t \to t_0 \tag{4.2}$$

where A is some positive constant.

It turns out [**10, 12**] for the general N-body problem that if there is a collision of any kind at $t = t_0$, then the colliding particles must approach each other like $(t_0 - t)^{2/3}$; i.e., the rate of approach for collinear collisions extends to all possible kinds of collisions. This is important information about collisions, but it doesn't explain the complex nature of this singularity. Is it an algebraic branch point? Is it a logarithmic singularity? As indicated next, binary collisions always correspond to algebraic branch points. This is true even if several different binary collisions occur at the same time [**12**].

One way to show this for the collinear problem is to "blow-up" the singularity by defining $X(t)$ as $X(t)\,(t_0 - t)^{2/3} = x(t)$. To simplify the equations, assume that the time of collision is at $t_0 = 0$ (this just defines the origin on the time axis) and that we approach the collision through positive values. (The system is time reversible; the equations of motion are invariant with respect to the change of independent variable $-t$.) By substituting $[t^{2/3}X(t)]''$ into the defining equations of motion, the equations for $X(t)$ become

$$t^2X'' + (4/3)tX' - (2/9)X = -X^{-2}. \tag{4.3}$$

To convert this Euler differential equation into a form that admits analytic solutions, use the change of the independent variable $s = t^{1/3}$ to obtain

$$s^2X'' + 2sX' - 2X = -9X^{-2}, \tag{4.4}$$

where the primes now denote differentiation with respect to s. By using series methods from elementary ordinary differential equations and a standard majoriant argument to justify convergence, it follows that this system has an analytic solution in s. Moreover, in a neighborhood about the time of collision the solution is

$$x(t) = \sum_k a_k t^{2k/3}. \tag{4.5}$$

It follows from eq. 4.5 that *a binary collision for the system is an algebraic branch point* where $s = t^{1/3}$ serves as a local uniformizing variable. As t passes through zero, the solution runs through the sheets of the Reimann surface to emerge as though the collision corresponds to an exact rebound or a perfectly elastic collision. Consequently, we should expect that this collision singularity can be removed, and that it doesn't constitute a serious obstacle to the analysis of the system. (The same assertion does not hold for triple or more complicated collisions. Here, some of the exponents in the expansion from eq. 4.3 or 4.4 depend continuously on the value of the masses of the colliding particles. Thus, logarithmic singularities result. See [**14**] for triple collisions and [**12**] for a general discussion.)

Incidentally, eq. 4.5 provides the basis for an "intuitive" argument to explain why Sundman's change of independent variable, $ds = dt/r(t)$, works. As in our derivation of eq. 4.5, if we know that the time of the collision is t_0, then we could substitute $s = (t - t_0)^{1/3}$, or $ds = dt/3(t - t_0)^{2/3}$, to obtain an analytic solution. But, we don't know, a priori, when or even whether a collision will occur. However, perhaps one way to avoid this complication about when a collision occurs is to base the time change on the appropriate power of $r(t)$. After all, a collision occurs if and only if r approaches zero, so the growth properties of r determine when a collision happens. If such an approach is to be successful, then the growth properties of r need to be further exploited by using the appropriate choice of α so that r^α effectively replaces the $(t - t_0)^{2/3}$ term in the change of the independent variable. The asymptotic expression given in eq. 4.2 shows that $\alpha = 1$, so a natural choice for the change of variables is $ds = dt/r(t)$. This is the Sundman change of the independent variable.

After Sundman eliminated binary collisions from the equations of motion, the only remaining, real singularity for the transformed three-body problem is a triple collision. Triple collisions can be avoided by appealing to a result proved by Sundman, and already known to Weierstrass (see [**9**, p. 66]). This result uses the integral of angular momentum $\sum_i m_i \mathbf{r}_i x \mathbf{v}_i = \mathbf{c}$ where $\mathbf{c}$ is a vector constant of integration and where m_i, $\mathbf{r}_i$, and $\mathbf{v}_i$ are, respectively, the mass, the position vector, and the velocity vector of the ith particle. The Weierstrass-Sundman theorem asserts that if $\mathbf{c} \neq \mathbf{0}$, then the system cannot have a complete collapse. (For an elegant proof, see [**9**, p. 66].) Namely, if $N = 3$, then triple collisions cannot occur off of the algebraic variety $\mathbf{c} = \mathbf{0}$.

By restricting attention to $\mathbf{c} \neq \mathbf{0}$, we've removed all real singularities from the transformed equations for the three-body problem. One last constraint on the radius of convergence of a series solution is the possibility of complex-valued singularities;

are there imaginary collisions of the particles at imaginary values of time? There are. To show this and to develop some insight about how they are related to real behavior, consider elliptical solutions of the central force problem eq. 3.1. The solution $r(t) = (\mathbf{r}(t), \mathbf{r}(t))^{1/2}$ does not have a closed form representation. Instead, it is defined implicitly through the *Kepler equations*

$$r(u) = a(1 - \varepsilon \cos u) \qquad t = u - \varepsilon \sin u, \tag{4.6}$$

where ε is the eccentricity of the ellipse, a is the length of the semimajor axis, and u is a variable.

Elliptical solutions are well behaved without any possibility of a collision, so one might guess that the power series converges for all values of time. It doesn't; this is because there are complex singularities. By letting $u = u_1 + iu_2$ and $t = t_1 + it_2$, a computation that just involves the expansion of sine and cosine for complex values proves that there is an "imaginary collision" where $r = 0$ at

$$u_k = 2k\pi \pm i \cosh^{-1}(1/\varepsilon), \qquad k = 1, 2, \ldots, \qquad t_k = u_k - e \sin u_k. \tag{4.7}$$

Thus, not only do complex singularities occur, but $\mathrm{Re}(t_k)$ is determined precisely where, on the real axis, $r(t)$ assumes its minimum value. Moreover, the smaller the value of this minimum (i.e., the larger the value of ε), the closer this complex singularity is to the real axis. This suggests that, if in some manner, the three-body system can be forced to be "bounded" away from a triple collision, then perhaps the complex singularities are bounded away from the real axis.

This is what Sundman proved. His first step, which is a significant improvement over the Weierstrass-Sundman Theorem, was to show that if $\mathbf{c} \neq \mathbf{0}$, then for each value of t, $\max\{|\mathbf{r}_i(t) - \mathbf{r}_j(t)|\} > D(\mathbf{c}) > 0$. In other words, if $\mathbf{c} \neq \mathbf{0}$, then not only can there be no triple collision, but the system is strictly bounded away from a triple collision. (Subsequently, Sundman's result has been extended to the N-body problem for all values of N. See [5].) Using this fact along with the Cauchy existence theorem for differential equations, Sundman proved that the system has no complex singularities in a strip (depending on the value of $\mathbf{c}$) in the complex plane centered around the real axis. With this, the remainder of his proof is immediate. All one needs to do is to conformally map this strip to the unit disk. For instance, if the strip is $|\mathrm{Im}\, s| \leqslant \beta$, then the change of independent variables is $\sigma = (e^{\pi s/2\beta} - 1)/(e^{\pi s/2\beta} + 1)$. In this new system, the equations of motion have no singularities in the unit disk, so the resulting power series converges. Along the real axis in the unit disk, σ corresponds, in a 1–1 fashion, to all real values of time. Thus, in this way, Sundman proved the existence of power series solutions for the three-body problem that converges for all real values of time.

Sundman solved the three-body problem but, unfortunately, the series solution is of little practical value because it converges too slowly and it requires too many terms to achieve any interesting degree of accuracy. It is fairly easy to see why this is so. Remember, Sundman used two changes of the independent variable. The first was to eliminate the singularities due to binary collisions. As we see from eq. 4.2, this involves a change that has the form $s = (t_0 - t)^{1/3}$ near a binary collision. This change of variables "slows down" the dynamics with the accompanying effect that a numerical value of s tends to be larger than the corresponding numerical value of t. (This is similar to the situation in numerical integration where smaller step sizes are

used to achieve accuracy with rapid changes in the dynamics. Accompanying the smaller steps is an increase in the number of steps; this number of steps corresponds to the value of s.) Next, this larger value of s is exponentially mapped to the unit disk in the complex σ-plane. Thus, it follows that fairly small values of t can be identified with values of σ near the boundary of the unit disk. When this happens, we should expect the convergence to be slow. This occurs. In other words, while Sundman successfully solved the three-body problem, the theoretical properties of his analysis demonstrate that, in general, an universal power series solution of the three-body problem is impractical. His success in solving an age-old problem killed interest in this line of inquiry. Nevertheless, many of Sundman's results used to achieve his series solution continue to play a significant role in the development of celestial mechanics.

5. A footnote. The emphasis in this brief description of aspects of the Newtonian N-body problem has been on the role played by complex variables. It is appropriate to end with a historical aside about how certain problems from celestial mechanics played a role in the development of some of the classical results in complex variables.

As mentioned in Section 3, the solution for the simple, central force problem, $r(t)$, given by eq. 4.6 for the elliptic case, does not have a closed form representation. This solution is only defined implicitly through Kepler's equations. Consequently, several mathematicians sought to find a useful direct expression, even a power series solution, for r. A. Wintner [**17**], in his book *Analytical Foundations of Celestial Mechanics*, notes in a footnote on page 217, "*The direct proof of* [*this aspect of Kepler's equation*] *played an important historical role in the theory of analytic functions*... [*A*] *principal impetus for Cauchy's discoveries in complex function theory was his desire to find a satisfactory treatment for Lagrange's series* [*for Kepler's equation*]. *Cauchy was led to his fundamental theorem connecting the radius of convergence with the location of the nearest singularity, as well as to his maximum principle, precisely in his papers dealing* [*with this problem*]. *Also the facts usually referred to as the argument principle and Rouche's theorem were first observed in connection with this problem concerning Kepler's equation.*"

Acknowledgement. Some of the material described in this essay is known to experts in celestial mechanics (e.g., see S. Sternberg's *Celestial Mechanics*, Part I, Benjamin, 1969), but it doesn't seem to be well known outside of this community. I've used aspects of this material in my courses on celestial mechanics, on complex variables, and even in calculus. Indeed, this article is based on my introductory lecture for a "short course" on the Newtonian N-body problem I gave at Instituto Matematica Pura e Aplicada, Rio de Janeiro, Brazil during January and February of 1987. I would like to thank the institute and my host, J. Palis, Jr., for their hospitality. This paper was written with partial support from an NSF grant.

REFERENCES

1. R. Devaney, Blowing up singularities in classical mechanical systems, this MONTHLY, 89 (1982), 535–552.
2. P. Kustaanheimo, Spinor regularization of the Kepler motion, *Ann. Univ. Turku, Ser.* AI. 73 (1964).
3. P. Kustaanheimo and E. Stiefel, Perturbation theory of Kepler motion based on spinor regularization, *J. Reine Angew. Math.*, 218, (1965) 204–219.
4. T. Levi-Civita, Sur la regularisation du probleme des troi corps, *Acta Math.*, 42 (1920) 99–144.

5. C. Marchal and D. G. Saari, On the final evolution of the N-body problem, *J. Diff. Eqs.*, 20 (1976) 150–186.
6. J. Milnor, On the geometry of the Kepler problem, this MONTHLY, 90 (1983) 353–364.
7. J. Moser, Regularization of Kepler's problem and the averaging method on a manifold, *Comm. Pure Appl. Math.*, 23 (1970) 609–636.
8. R. McGehee, Triple collisions in the collinear three body problem, *Invent. Math.*, 27 (1974) 191–227.
9. Harry Pollard, Celestial Mechanics, Carus Math Monograph # 18, MAA, 1976.
10. H. Pollard and D. G. Saari, Singularities of the N-body problem, I, *Arch. Rat. Mech. Anal.*, 30 (1968) 263–269.
11. D. G. Saari, Improbability of collisions in Newtonian gravitational systems, *Trans. Amer. Math. Soc.*, 162 (1971), 267–271; Paper II, *Trans. Amer. Math. Soc.*, 181 (1973) 351–368.
12. _____, Manifold structure for collisions and for hyperbolic-parabolic orbits in the N-body problem, *J. Diff. Eqs.*, 55 (1984) 300–329.
13. _____, A global existence theorem for the four body problem of Newtonian mechanics, *J. Diff. Eqs.*, 26 (1977) 80–111.
14. C. L. Siegel, Der Dreierstross, *Ann. Math.*, 42 (1941) 127–168.
15. E. L. Stiefel, and G. Scheifel, Linear and Regular Celestial Mechanics, Springer-Verlag, New York, 1971.
16. K. F. Sundman, Recherches sur le probleme des trois corps, *Acta Soc. Sci. Fennicae*, 34 no. 34 (1907).
17. A. Wintner, The Analytical Foundations of Celestial Mechanics, Princeton University Press, 1941.

The Square Pyramid Puzzle

W. S. ANGLIN, *University of Notre Dame*, Notre Dame, IN 46556

BILL ANGLIN received an interdisciplinary Ph.D. in Mathematics and Philosophy at McGill University in 1987. He is currently a Postdoctoral Fellow in Philosophy of Religion at Notre Dame University, and he is also continuing his work in classical Number Theory.



In 1875 Edouard Lucas challenged the readers of the *Nouvelles Annales de Mathématiques* to prove the following:

> A square pyramid of cannon-balls contains a square number of cannon-balls only when it has 24 cannon-balls along its base [**4**].

In other words, the only nontrivial, natural number solution of

$$1^2 + 2^2 + 3^2 + \cdots + x^2 = y^2$$

is $x = 24$ and $y = 70$.

We begin this article by recalling the history of this Diophantine equation. We then present a new, elementary solution of it.

In 1876 Moret-Blanc gave a "proof" of Lucas's assertion which, although incomplete, at least gives a good framework for solving it [**7**]. Moret-Blanc divided the problem into two cases, one with x even and one with x odd. This is exactly the approach we take in this paper. In 1877, having pointed out that there was a gap in Moret-Blanc's argument, Lucas published a proof which also contained a gap [**5**]. Lucas could handle the case with x even but not the case with x odd. In 1918 G. N. Watson managed to fill Lucas's gap with about 14 pages of a specially extended theory of Jacobian elliptic functions [**9**]. This was the first complete proof of Lucas's assertion but it was very complicated. In 1952 Ljunggren used Skolem's method to give a simpler, more arithmetic solution to the square pyramid problem [**3**]. In 1966 Baker and Davenport used a theorem from transcendental number theory to solve the simultaneous Diophantine equations $3x^2 - 2 = y^2$ and $8x - 7 = z^2$ [**1**]. Their method, recently made more practical by Waldschmidt, can easily be adapted to demonstrate Lucas's proposition. In 1975 Kanagasabapathy and Ponnudurai gave an elementary solution for the simultaneous Diophantine equations $3x^2 - 2 = y^2$ and $8x^2 - 7 = z^2$ [**2**], and, in 1985, De Gang Ma used the ideas of Kanagasabapathy and Ponnudurai to arrive at an elementary solution to Lucas's problem [**6**]. This was the first complete solution accessible at the undergraduate level.

In this paper we draw on one of Ma's ideas to give a new proof of Lucas's assertion. This proof is so short and simple that it is surprising that it escaped mathematicians for over a hundred years.

I

Our equation is equivalent to $x(x+1)(2x+1) = 6y^2$. In this section we solve it assuming that x is even, and, in the next section, we solve it assuming that x is odd. Suppose, then, that x is even. Under this assumption we can solve the equation with the help of the following three lemmas which were probably all known to Fermat.

LEMMA 1. *The area of a Pythagorean triangle is never a square.*

Proof. Suppose, on the contrary, there are Pythagorean triangles with square areas. Let w^2 be the smallest area for which such a triangle exists. Let x and y be the legs of a Pythagorean triangle with area w^2. Then $x^2 + y^2 = z^2$ for some integer z, and $xy/2 = w^2$. Since w is minimal, x and y are relatively prime and, without loss of generality, we may take x odd and y even. (It follows by a congruence consideration modulo 4 that x and y are not both odd.) By the well-known theorem for Pythagorean triangles, there are relatively prime positive integers r and s of different parity such that $x = r^2 - s^2$ and $y = 2rs$. Hence $(r^2 - s^2)rs = w^2$ and $s/4 < s \leqslant w^2$. Since r, s, $r - s$ and $r + s$ are pairwise relatively prime, and since $(r - s)(r + s)rs = w^2$, it follows that there are positive integers a, b, c and d such that $r = a^2$, $s = b^2$, $a^2 - b^2 = r - s = c^2$ and $a^2 + b^2 = r + s = d^2$. Note that c and d are relatively prime since $r - s$ and $r + s$ are relatively prime. Noting also that c and d are odd (because r and s have different parity), let $X = (c + d)/2$ and $Y = (d - c)/2$. Then X and Y are relatively prime and $X^2 + Y^2 = a^2$. Hence one of X and Y is even, and $XY/2 = (d^2 - c^2)/8 = b^2/4 = s/4$ is a square integer. Since the triangle with sides X, Y and a is a Pythagorean triangle with square area $s/4$, it follows from the minimality of w^2 that $w^2 \leqslant s/4$. Contradiction.

LEMMA 2. *There are no positive integers x such that $2x^4 + 1$ is a square.*

Proof. To obtain a contradiction, suppose that (x, y) is the least positive integer solution of $2x^4 + 1 = y^2$. Then for some positive integer s, $y = 2s + 1$ and $x^4 = 2s(s + 1)$. If s is odd then s and $2(s + 1)$ are relatively prime and, for some integers u and v, $s = u^4$ and $2(s + 1) = v^4$. This gives $2(u^4 + 1) = v^4$ with u odd and v even. Hence we have $2(1 + 1) \equiv 0 \pmod 8$. Since this is impossible, s cannot be odd. Since s is even, $2s$ and $s + 1$ are relatively prime, and there are integers u and v, both greater than 1, such that $2s = u^4$ and $s + 1 = v^4$. Let w be the positive integer such that $u = 2w$. Let a be the positive integer such that $v^2 = 2a + 1$. Then $u^4/2 + 1 = s + 1 = v^4$ so that $2w^4 = (v^4 - 1)/4 = a(a + 1)$. Since $v^2 = 2a + 1$, it follows from congruence considerations modulo 4 that a is even. Since $2w^4 = a(a + 1)$, it follows that there are positive integers b and c such that $a = 2b^4$ and $a + 1 = c^4$. However, this implies that $2b^4 + 1 = (c^2)^2$ and hence $y \leqslant c^2$ (by the minimality of (x, y)). On the other hand, $c^2 \leqslant a + 1 < v^2 \leqslant s + 1 < y$. Contradiction.

LEMMA 3. *There is exactly one positive integer x, namely, 1, such that $8x^4 + 1$ is a square.*

Proof. Suppose $8x^4 + 1 = (2s + 1)^2$. Then $2x^4 = s(s + 1)$. If s is even then there are integers u and v such that $s = 2u^4$ and $s + 1 = v^4$. In that case, $2u^4 + 1 = s + 1 = v^4$ and, by Lemma 2, $u = 0$ and, hence, $x = 0$. If s is odd

then there are integers u and v such that $s = u^4$ and $s + 1 = 2v^4$. In that case, $u^4 + 1 = 2v^4$. Since u is odd, a congruence consideration modulo 4 shows that v is odd. Squaring both sides of $u^4 + 1 = 2v^4$, we obtain $4v^8 - 4u^4 = u^8 - 2u^4 + 1$ and hence $(v^4 - u^2)(v^4 + u^2) = ((u^4 - 1)/2)^2$, an integer square. Since v^4 and u^2 are relatively prime, it follows that both $(v^4 - u^2)/2$ and $(v^4 + u^2)/2$ are integer squares. Now $(v^2 - u)^2 + (v^2 + u)^2 = 4(v^4 + u^2)/2 = A^2$ and $(v^2 - u)(v^2 + u)/2 = (v^4 - u^2)/2 = B^2$. By Lemma 1, this is impossible unless $v^2 = \pm u$. Since $u^4 + 1 = 2v^4$, we obtain $u^4 - 2u^2 + 1 = 0$ and $u^2 = 1$. From this it follows that $s = 1$ and $x = \pm 1$.

With the above three lemmas, we are now in a position to solve $x(x + 1)(2x + 1) = 6y^2$ under the assumption that x is even. Suppose, then, that x is even. Then $x + 1$ is odd. Since x, $x + 1$ and $2x + 1$ are relatively prime in pairs, it follows that $x + 1$ and $2x + 1$ (both being odd) are either squares or triples of squares. Thus $x + 1 \not\equiv 2 \pmod 3$ and $2x + 1 \not\equiv 2 \pmod 3$. Hence $x \equiv 0 \pmod 3$, and for some nonnegative integers p, q and r, we have $x = 6q^2$, $x + 1 = p^2$ and $2x + 1 = r^2$. Thus $6q^2 = (r - p)(r + p)$. Since p and r are both odd, 4 is a factor of $(r - p)(r + p) = 6q^2$ and thus q is even. Let q' be the integer such that $q = 2q'$. We now have $6q'^2 = ((r - p)/2)((r + p)/2)$ and, since $(r - p)/2$ and $(r + p)/2$ are relatively prime (because r^2 and p^2 are relatively prime), we obtain one of the following two cases.

Case (i). One of $(r - p)/2$ and $(r + p)/2$ has the form $6A^2$ and the other has the form B^2 (where A and B are nonnegative integers). Then $p = \pm(6A^2 - B^2)$ and $q = 2AB$. Since $6q^2 + 1 = x + 1 = p^2$, we have $24A^2B^2 + 1 = (6A^2 - B^2)^2$ or $(6A^2 - 3B^2)^2 - 8B^4 = 1$. By Lemma 3, $B = 0$ or 1 and, hence, $x = 6q^2 = 0$ or 24. The only nontrivial solution is thus with $x = 24$.

Case (ii). One of $(r - p)/2$ and $(r + p)/2$ has the form $3A^2$ and the other has the form $2B^2$ (where A and B are nonnegative integers). Then $p = \pm(3A^2 - 2B^2)$ and $q = 2AB$. This gives $24A^2B^2 + 1 = (3A^2 - 2B^2)^2$ and hence $(3A^2 - 6B^2)^2 - 2(2B)^4 = 1$. By Lemma 2, $B = 0$ and hence $x = 6q^2 = 0$.

Thus when x is even, the only solution to Lucas's puzzle is $x = 24$ cannonballs along the base of the square pyramid.

II

We have solved Lucas's problem under the assumption that x is even. In this section we solve it under the assumption that x is odd. To do this, we first investigate the solutions of the Diophantine equation $X^2 - 3Y^2 = 1$.

Let $a = 2 + \sqrt{3}$ and $b = 2 - \sqrt{3}$. Note that $ab = 1$. Where n is any nonnegative integer, let $u_n = (a^n + b^n)/2$ and $v_n = (a^n - b^n)/(2\sqrt{3})$. Then u_n and v_n are integers, and it is a well-known result (from the theory of the Pell equation) that when $n \geq 1$, (u_n, v_n) is the nth positive integer solution of $X^2 - 3Y^2 = 1$. (See [**8**, pp. 209–211].) Of course, when $n = 0$, we have the solution $X = 1$ and $Y = 0$.

In order to show that $x = 1$ is the only odd positive integer such that $x(x + 1)(2x + 1)$ has the form $6y^2$, we use the following lemmas.

LEMMA 4. *Where m and n are nonnegative integers, $u_{m+n} = u_m u_n + 3v_m v_n$ and $v_{m+n} = u_m v_n + u_n v_m$. Also if $m - n \geq 0$, then $u_{m-n} = u_m u_n - 3v_m v_n$ and $v_{m-n} = -u_m v_n + u_n v_m$.*

Proof. This follows by straight calculation from the definitions of u_n and v_n.

Using Lemma 4, it is not hard to obtain the following result.

LEMMA 5. *Where m is a nonnegative integer,* $u_{m+2} = 2u_1u_{m+1} - u_m$ *and* $v_{m+2} = 2u_1v_{m+1} - v_m$.

Using the fact that (u_m, v_m) is a solution of $X^2 - 3Y^2 = 1$, we also have the following.

LEMMA 6. *Where m is a nonnegative integer,* $u_{2m} = 2u_m^2 - 1 = 6v_m^2 + 1$ *and* $v_{2m} = 2u_mv_m$.

LEMMA 7. *Let m, n and r be nonnegative integers such that* $2rm - n$ *is nonnegative. Then* $u_{2rm \pm n} \equiv (-1)^r u_n \pmod{u_m}$.

Proof. Using mathematical induction on r together with Lemma 4, we can show that $u_{(2r+1)m} \equiv 0 \pmod{u_m}$ and $v_{2rm} \equiv 0 \pmod{u_m}$. Since, by Lemma 6, $u_{2rm} = 2u_{rm}^2 - 1 = 6v_{rm}^2 + 1$, it follows that $u_{2rm} \equiv (-1)^r \pmod{u_m}$. Thus $u_{2rm \pm n} = u_{2rm}u_n \pm 3v_{2rm}v_n \equiv (-1)^r u_n \pmod{u_m}$ (using Lemma 4).

Let us consider the first few values of u_n. Starting with $n = 0$, we have 1, 2, 7, 26, 97, 362, and so on. If we consider these values modulo 5, we have 1, 2, 2, 1, 2, 2, ... By Lemma 5 it follows that this sequence is periodic. If we consider the values of u_n modulo 8, we obtain 1, 2, 7, 2, 1, 2, ... By Lemma 5, this is a purely periodic sequence with period length 4. Note that when n is even, u_n is odd. Using the laws of quadratic reciprocity, the above comments lead us to the following two lemmas.

LEMMA 8. *If n is even then* u_n *is an odd nonmultiple of* 5 *and* $\left(\frac{5}{u_n}\right) = 1$ *iff n is a multiple of* 3.

LEMMA 9. *If n is even then* u_n *is odd and* $\left(\frac{-2}{u_n}\right) = 1$ *iff n is a multiple of* 4.

The following and final lemma was first proved by Ma.

LEMMA 10. *Where n is a nonnegative integer,* u_n *has the form* $4M^2 + 3$ *only when* $u_n = 7$.

Proof. Suppose $u_n = 4M^2 + 3$. Then $u_n \equiv 3$ or $7 \pmod 8$ and, from the sequence of values of u_n modulo 8, it follows that n has the form $8k \pm 2$. Suppose $n \neq 2$ (and hence $u_n \neq 7$). Then we can write n in the form $2k'2^s \pm 2$ where k' is odd and $s \geqslant 2$. By Lemma 7, $u_n = u_{2k'2^s \pm 2} \equiv (-1)^{k'} u_2 \pmod{u_{2^s}}$. Since k' is odd and $u_2 = 7$, it follows that $4M^2 = u_n - 3 \equiv -10 \pmod{u_{2^s}}$ and, hence,

$$\left(\frac{-2}{u_{2^s}}\right)\left(\frac{5}{u_{2^s}}\right) = \left(\frac{-10}{u_{2^s}}\right) = \left(\frac{4M^2}{u_{2^s}}\right) = 1.$$

By Lemma 9 it follows that the first factor on the left is 1. By Lemma 8 it follows that the second factor on the left is -1. Since this is impossible, we may conclude that $n = 2$ and hence $u_n = 7$.

Suppose now that x is an odd positive integer and $x(x+1)(2x+1) = 6y^2$ for some integer y. Since x, $x+1$ and $2x+1$ are relatively prime in pairs, it follows that x is either a square or a triple of a square, and hence $x \not\equiv 2 \pmod 3$. Moreover, $x+1$, being even, is either double a square or six times a square, and, hence, $x+1 \not\equiv 1 \pmod 3$. Thus $x \equiv 1 \pmod 3$ and hence $x+1 \equiv 2 \pmod 3$ and $2x+1 \equiv 0 \pmod 3$. Thus for some nonnegative integers u, v and w, we have $x = u^2$, $x+1 = 2v^2$ and $2x+1 = 3w^2$. From this we obtain $6w^2+1 = 4x+3 = 4u^2+3$. Also $(6w^2+1)^2 - 3(4vw)^2 = 12w^2(3w^2+1-4v^2)+1 = 12w^2(2x+1+1-2(x+1))+1 = 1$. Hence, by Lemma 10, $6w^2+1 = 7$. Thus $w = 1$ and $x = 1$. This gives us the trivial 1 cannonball solution to Lucas's problem.

We may conclude that if a square number of cannonballs are stacked in a square pyramid then there are exactly 4900 of them.

I would like to thank Professor J. Lambek for his kind help and encouragement.

REFERENCES

1. A. Baker and H. Davenport, The Equations $3x^2 - 2 = y^2$ and $8x^2 - 7 = z^2$, *Quarterly Journal of Mathematics*, ser. 2, 20 (1969) 129–37.
2. P. Kanagasabapathy and T. Ponnudurai, The simultaneous Diophantine equations $y^2 - 3x^2 = -2$ and $z^2 - 8x^2 = -7$, *Quarterly Journal of Mathematics*, ser. 2, 26 (1975) 275–78.
3. W. Ljunggren, New solution of a problem proposed by E. Lucas, *Norsk Mat. Tidsskrift*, 34 (1952) 65–72.
4. Edouard Lucas, Question 1180, *Nouvelles Annales de Mathématiques*, ser. 2, 14 (1875) 336.
5. ——, Question 1180, *Nouvelles Annales de Mathématiques*, ser. 2, 16 (1877) 429–32.
6. De Gang Ma, An elementary proof of the solution to the Diophantine equation $6y^2 = x(x+1)(2x+1)$, *Sichuan Daxue Xuebao*, 4 (1985) 107–16.
7. M. Moret-Blanc, Question 1180, *Nouvelles Annales de Mathématiques*, ser. 2, 15 (1876) 46–48.
8. I. Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, 4th ed., John Wiley and Sons, New York, 1980.
9. G. N. Watson, The problem of the square pyramid, *Messenger of Mathematics*, 48 (1918–19) 1–22.

Editor's Corner: The Euclidean Algorithm Strikes Again

STAN WAGON
Department of Mathematics, Smith College, Northampton, MA 01063

On page 144 of this issue of the MONTHLY Don Zagier presents an extremely short, elegant, and elementary proof of the classic result that any prime p congruent to 1 (mod 4) is a sum of two squares. The problem of *finding* a representation of p as $\alpha^2 + \beta^2$ often arises in computational number theory; for example, it is a key step in factoring Gaussian integers into prime Gaussian integers. Fortunately, this problem can be solved by a very fast and easy-to-program algorithm—it is not widely known, although it uses nothing beyond undergraduate mathematics. In this note we will present the algorithm and a proof of its correctness.

We pause for a moment to point out a beautiful formula for α and β due to Gauss: If $p = 4k + 1$ then $\alpha = \left\langle \frac{1}{2}\binom{2k}{k} \right\rangle$ and $\beta = \langle (2k)!\alpha \rangle$ are such that $p = \alpha^2 + \beta^2$ (here $\langle n \rangle$ denotes the residue of n mod p with $|\langle n \rangle| < \frac{p}{2}$). For a proof see [**3**, chaps. 4–5]. Simple as it is, Gauss's formula does not seem to be of any value for computation of α and β.

Now, let p be a prime of the form $4k + 1$. The algorithm has two steps:

A. Find x such that $x^2 \equiv -1 \pmod{p}$.

B. Apply the Euclidean algorithm to p and x; the first two remainders that are less than $\sqrt{p}$ may be taken as α and β.

Step A is easy in practice: First find c, a quadratic nonresidue of p; recalling Euler's criterion (which states that for a not divisible by p, a is (resp., is not) a quadratic residue mod p iff $a^{(p-1)/2} \equiv +1$ (resp., -1) mod p), we have that $c^{2k} \equiv -1 \pmod{p}$, whence x may be taken to be c^k mod p. How to find a nonresidue? Start with $c = 2$ and let c take on increasingly larger prime values, checking each until a nonresidue is found. To do this efficiently, recall that 2 is a nonresidue mod p iff $p \equiv 5 \pmod{8}$ and 3 is a nonresidue mod p iff $p \equiv 2 \pmod{3}$. For primes $c > 3$, observe first that, because $p \equiv 1 \pmod{4}$, quadratic reciprocity implies that c is a nonresidue mod p iff p is a nonresidue mod c iff d is a nonresidue mod c, where d is the mod-c reduction of p. Now, one can simply use Euler's criterion via $d^{(c-1)/2}$ mod c. Alternatively, continue to reduce via quadratic reciprocity and the Jacobi symbol (see, e.g. [**9**, §3.3]).

The search for a nonresidue cannot go on too long, at least if, as expected, the Extended Riemann Hypothesis is true: Under ERH the first nonresidue is less than $2(\log p)^2$ [**1**] (and hence this part of the algorithm runs in polynomial time). In practice things are generally much better than the ERH bound. For example (see [**1**]), if $p \equiv 1 \pmod{4}$ and $p < 1{,}000{,}000$, then the least nonresidue is at most 37, while $2(\log p)^2$ is about 381 for primes near 1,000,000. As an experiment, I computed the least nonresidue for the first 418 primes[1] greater than 10^{15} (for which

[1]Actually, the 418 subjects of the experiment are only probably prime. The PrimeQuery function in *Mathematica* is highly reliable but has not been proved to be failsafe, and it is easy to forget that some of

the ERH-bound is 2,385). The first nonresidue is never greater than 29, and only 19 of the primes have their first nonresidue outside of $\{2, 3, 5, 7\}$.

The essence of this two-squares algorithm was first presented by Serret and Hermite in 1848 [**8**, **11**] using the language of continued fractions. The streamlined form presented above is due to G. Cornacchia [**3**] (see also [**2**]). Here we present a self-contained proof using only elementary facts about the Euclidean algorithm.

First, an example. Let $p = 848654483879497562821$. Because $p \equiv 5 \pmod 8$, c may be taken to be 2 and x is easily calculated by the power-mod algorithm. Applying the Euclidean algorithm yields the following remainders:

$p =$ 848654483879497562821	25504875423829707
$x =$ 354060813206257083018	3100076540181815
848654483879497562821	704263102375187
140532857466983396785	283024130681067
72995098272290289448	138214841013053
67537759194693107337	6594448654961
5457339077597182111	6325867913833
2049690263526922005	268580741128
1357958550543338101	148510867889
691731712983583904	120069873239
666226837559754197	$\alpha =$ 28440994650
	$\beta =$ 6305894639

And, as predicted, $28440994650^2 + 6305894639^2 = 848654483879497562821$. This computation took a quarter of a second using *Mathematica* on a Macintosh.

Now to the proof that the algorithm works (which also provides a proof of existence of α and β). Some notation: if a and b are the starting values of the Euclidean algorithm, let the sequence of remainders be $r_0\ (= a)$, $r_1\ (= b)$, $r_2, \ldots, r_n$, with $r_n = \gcd(a, b)$. Let q_i denote the quotients, that is, $q_1 = \lfloor a/b \rfloor$, $q_2 = \lfloor b/r_2 \rfloor, \ldots$. Recall that along with the remainders and quotients, one often computes two sequences $\{s_i\}$ and $\{t_i\}$ so that $r_i = s_i a + t_i b$, $i = 0, 1, \ldots, n$. We will need only the ts, which are defined as follows: $t_0 = 0$, $t_1 = 1$, and $t_{i+1} = t_{i-1} - q_i t_i$. It is easy to see by induction that $t_i b \equiv r_i \pmod a$ for each i (this forward computation of t—and of s, which is defined similarly—is much easier than the "back-substitution" method discussed in some texts). We need some additional facts about the t-sequence. It is immediate from the definition that these integers alternate in sign and that the sequence $\{|t_i| : i = 0, 1, \ldots, n\}$ is strictly increasing. We will need one more term than is usual, so let t_{n+1} be $t_{n-1} - q_n t_n$. Because $|t_{i+1}| = q_i |t_i| + |t_{i-1}|$ and $|t_{i-1}| < |t_i|$, the sequence $\{|t_i| : i = n + 1, n, \ldots, 1\}$ is a Euclidean algorithm remainder sequence starting from the two values $|t_{n+1}|$ and $|t_n|$.

Let's look at an example more closely, where a is p, the prime in question, and b is x which, from now on, denotes the *smaller* of the two square roots of

the "primes" may be composite. While the experiment was running I thought it was checking actual primes. I was probably right. (*Added in proof*: Ilan Vardi has since confirmed, using an elliptic curve method, that the 418 "primes" are in fact prime.)

$-1 \bmod p$ (the other root is $p - x$). Here's an example with $p = 73$, for which $x = 27$ and $n = 6$:

r	q	t
73 ($p = r_0$)		0 (t_0)
27 ($x = r_1$)	2 (q_1)	1
19	1	-2
8	2	3
3	2	-8
2	1	19
1 (r_6)	2 (q_6)	-27 (t_6)
		73

The surprising thing is that the t-sequence, in absolute value, is simply the *reverse* of the remainder sequence generated by p and x. This is the key point in the proof that the algorithm works (in this case: $8^2 + 3^2 = 73$). To carry out the proof, we list the facts we need in the case of r-, q-, and t-sequences defined from a prime p (a prime congruent to 1 mod 4) and x (the smaller of the two square roots of $-1 \bmod p$).

1. The t-sequence is, in absolute value, just the reverse of the r-sequence.
2. n is even.
3. The q-sequence is symmetric about its center.
4. p divides each $r_i^2 + t_i^2$.
5. $r_{n/2}$ is the first remainder under $\sqrt{p}$.

The validity of the algorithm follows from these facts. By (4), p divides $s = r_{n/2}^2 + t_{n/2}^2$ which, by (1), equals $r_{n/2}^2 + r_{n/2+1}^2$. But by (5) each of these two remainders is under $\sqrt{p}$, so $s < 2p$; because s is divisible by p, s must in fact equal p.

Before proving (1)–(5) we need an additional result, one first discovered in the context of continued fractions. This result is valid for any starting values, so we revert momentarily to the general starting values a and b, with $\gcd(a,b) = 1$. The proof is a straightforward induction on n, combining the inductive assumptions that $b = f(q_2, \ldots, q_n)$ and $r_2 = f(q_3, \ldots, q_n)$ with the equation $a = q_1 b + r_2$.

EULER'S FORMULA. There is a function f, described below, such that $a = f(q_1, \ldots, q_n)$. The function f is a sum of products: First take the product of all the qs; then add every product that can be obtained by omitting any pair of consecutive terms; then add every product that can be obtained by omitting any two separate pairs of consecutive terms; and so on. For example: if $n = 5$, then $f(q_1, \ldots, q_5) = q_1q_2q_3q_4q_5 + q_3q_4q_5 + q_1q_4q_5 + q_1q_2q_5 + q_1q_2q_3 + q_5 + q_3 + q_1$. (The empty product, which occurs if n is even, yields the summand 1.)

We now return to the starting values p and x and conclude the proof of validity of the two-squares algorithm by proving (1)–(5).

1. First note that $t_{n+1} = \pm p$. This is because the t-sequence is, in absolute value, a Euclidean algorithm sequence with the same quotients, but in reverse order, as the r-sequence; the function in Euler's formula is invariant under reversal of its arguments, so $|t_{n+1}|$, being the same function of the quotients as p is, must equal p. Now, the fundamental congruence $t_i b \equiv r_i \pmod{a}$ implies that $t_n x \equiv 1 \pmod{p}$. Multiply this congruence by x to get $t_n \equiv -x$; because $|t_n| < p$, t_n must be either $-x$ or $p - x$. But the latter is impossible, for then the t-sequence (reversed, and in absolute value) would be a remainder sequence beginning: $p, p - x, x, \ldots$, which would mean it would be longer than the r-sequence, contradiction. Therefore $t_n = -x$. Now, because the $|t|$-sequence is, working backwards, a remainder sequence beginning with p and x, it coincides with the r-sequence.

2. This is immediate because the t-sequence alternates in sign and, as just proved, t_n is negative.

3. By (1), the quotients in the $|t|$-sequence (working back from t_{n+1}) are identical to the elements of the q-sequence; but from the equation $|t_{i+1}| = |t_{i-1}| + q_i|t_i|$, these quotients are the reverse of those in the q-sequence. Therefore each $q_i = q_{n+1-i}$.

4. Simply square both sides of the congruence $t_i x \equiv r_i \pmod{p}$, using $x^2 \equiv -1 \pmod{p}$.

5. Use the symmetry of the quotients and Euler's formula applied to the second half of the sequence of quotients to show that all the terms occurring in the square of $f\left(q\frac{n}{2} + 1, \ldots, q_n\right)$ also occur in the sequence $f(q_1, \ldots, q_n)$. That is, $r_{n/2}^2 = f(q_1, \ldots, q_{n/2}) f\left(q\frac{n}{2} + 1, \ldots, q_n\right)$ and each term in this expansion is also in the expansion of $f(q_1, \ldots, q_n) = p$. Thus $r_{n/2}^2 \leq p$. Since p is prime, the inequality is strict. It remains to show that the preceding remainder, $r_{n/2-1}$, squares to a value larger than p. This can be done in the same way by showing that every product in the f-representation of p occurs in the square of the f-representation of the preceding remainder, which is $f\left(q_1, \ldots, q\frac{n}{2} + 1\right) f\left(q_{n/2}, q\frac{n}{2} + 1, \ldots, q_n\right)$. The details are straightforward.

There does exist a provably polynomial-time algorithm for representing a prime as a sum of two squares, though in practice it is not as fast as the algorithm just given. It is not too hard to show that if α is such that $p + 2\alpha$ is the number of points (x, y) in $\mathbb{Z}_p \times \mathbb{Z}_p$ such that $y^2 = x^3 - x$, then α^2 is one of the two squares summing to p. (This can be done with the help of Jacobsthal's formula (see [**5**, p. 123]), which asserts that one of the two squares summing to p is one half of the sum over $\mathbb{Z}_p$ of the Legendre symbols $((x^3 - x)/p)$.) Now, there is a polynomial-time algorithm for finding the number of points on an elliptic curve (due to R. Schoof [**10**]) and it can be used to find $p + 2\alpha$, thus yielding a polynomial-time algorithm for finding α.

We close by mentioning some recent work on the more general problem of representing an integer as $f\alpha^2 + g\beta^2$ for given positive integers f and g (the case $f = 1$ was considered by Cornacchia [**4**]; [**12**] treats the case $f = 1$, $g = 5$). In [**7**] K. Hardy, J. B. Muskat, and K. S. Williams show that the Euclidean algorithm technique can be used, even when the number to be represented is composite. For

the two-squares problem ($f = g = 1$) their approach finds all *primitive solutions* (i.e., those with relatively prime α and β) as follows: First factor the given integer n; then for each prime dividing n find a square root of -1 modulo that prime (if one of these roots fails to exist—that is, the prime is congruent to 3 mod 4—then there is no primitive solution). Standard techniques (see, e.g. [**6**, p. 67]) can be used to transform these roots to a complete set of roots of -1 modulo each of the prime powers dividing n. These, in turn, can be used (with the help of the Chinese Remainder Theorem) to get all values of x such that $x^2 \equiv 1$ modulo n. Now, for each such x apply the Euclidean algorithm to n and the smaller of x, $n - x$. The first two remainders under $\sqrt{n}$ form a primitive solution, and all primitive solutions arise from the various roots x in this way. The proofs of facts 1–5 do not require that p be prime (that was used only in Step A), and so provide a proof that the preceding algorithm generating primitive solutions for composite n is valid. If *all* solutions to $n = \alpha^2 + \beta^2$ are sought, then find the primitive solutions for each integer of the form n/d^2 (where d^2 divides n), multiplying each solution by d to obtain a representation of n.

Acknowledgment. The author is grateful to Don Zagier for telling him of the algorithm presented in this note and suggesting improvements in the exposition.

REFERENCES

1. E. Bach, Analytic Methods in the Analysis and Design of Number-Theoretic Algorithms, Cambridge, MIT Press, 1985.
2. J. Brillhart, Note on representing a prime as a sum of two squares, *Math. Comp.* 26 (1972) 1011–1013.
3. S. Chowla, The Riemann Hypothesis and Hilbert's Tenth Problem, New York, Gordon and Breach, 1965.
4. G. Cornacchia, Su di un metodo per la risoluzione in numeri interi dell'equazione $\sum_{h=0}^{n} C_h x^{n-h} y^h = P$, *Giornale di Matematiche di Battaglini*, 46 (1908) 33–90.
5. H. Davenport, *The Higher Arithmetic*, 5th ed., Cambridge Univ. Pr., Cambridge, 1984.
6. D. Flath, Introduction to Number Theory, Wiley, New York, 1989.
7. K. Hardy, J. B. Muskat, and K. S. Williams, A deterministic algorithm for solving $n = fu^2 + gv^2$ in coprime intergers u and v, *Math. Comp.*, forthcoming.
8. C. Hermite, Note au sujet de l'article précédent, *J. Math. Pures Appl.*, 5 (1848) 15.
9. I. Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, second ed., New York, Wiley, 1966.
10. R. Schoof, Elliptic curves over finite fields and the computation of square roots mod p, *Math. Comp.*, 44 (1985) 483–494.
11. J. A. Serret, Sur un théorème rélatif aux nombres entières, *J. Math. Pures Appl.*, 5 (1848) 12–14.
12. P. Wilker, An efficient algorithmic solution of the Diophantine equation $u^2 + 5v^2 = m$, *Math. Comp.*, 35 (1976) 1347–1352.

LETTERS TO THE EDITOR

Editor:

The following comments on two recent articles in the MONTHLY may be of interest. Regarding the paper [1] on the Radon-Nikodým theorem, an essentially identical argument appears as one of the three proofs under the heading "First Proof (Probabilistic)" in the graduate text book [2], on pages 258–259. There it was called "Probabilistic" since the sequence, whose limit is the desired Radon-Nikodým derivative, forms a martingale with a directed index set. Similarly the following article [3] contains a short proof of the fact that, when $\nu \ll \mu$ are measures of which μ is finite and ν arbitrary, ν has a purely infinite set. This is also found (with even a shorter proof) on page 273 of [2].

REFERENCES

1. R. C. Bradley, An elementary treatment of the Radon-Nikodým Theorem, this MONTHLY, 96 (1989) 437–446.
2. M. M. Rao, Measure Theory and Integrations, Wiley-Interscience, New York, (1987).
3. A. Wilansky, On the proof of the Radon-Nikodým Theorem, this MONTHLY, 96 (1989) 441.

M. M. RAO
Mathematics Department
University of California
Riverside, CA 92521

Editor:

I wish to make some observations concerning the paper "On the proof of the Radon-Nikodým Theorem" by Albert Wilansky that appeared in the MONTHLY on p. 441 of vol. 96(5), May 1989. The main result of this paper is an exercise in [1]; see Exercise 31(7) on p. 131. This exercise refers to Exercise 30.11 which in turn refers to Exercise 17.3, and the latter exercise contains all the techniques used by the author. Halmos calls the technique the *method of exhaustion*.

The idea of the method of exhaustion is as follows: Suppose that μ and ν are nonnegative measures on a sigma algebra Σ of subsets of a set S and that the measure μ is finite. Define α to be the sup of the set of μ measures of sets that are ν sigma-finite. Choose an expanding sequence (A_n) of sets such that for each n, ν is sigma-finite on A_n and $\mu(A_n) > \alpha - 1/n$. Define

$$A = \bigcup_{n=1}^{\infty} A_n.$$

Then $\mu(A) = \alpha$, and for every measurable subset E of $S \setminus A$, either $\mu(E) = 0$ or $\nu(E) = \infty$.

In my opinion, this approach to the method of exhaustion is more elegant than the author's. The author also seems to suggest that his method is being offered as an alternative to a transfinite recursion; but I don't think that transfinite methods would normally be selected in this situation.

Finally, there is no need to assume $\nu \ll \mu$ for this argument to work. As a matter of fact, the argument is even more interesting when the absolute continuity goes the other way, i.e., $\mu \ll \nu$. See, for example, Theorem 2.2 in [2].

REFERENCES

1. Paul R. Halmos, Measure Theory, Van Nostrand, Princeton, NJ, 1950.
2. Jonathan Lewin and Mirit Lewin, A reformulation of the Radon-Nikodým theorem, *Proceedings of the AMS*, 47(2) (1975) 393–400.

JONATHAN LEWIN,
Kennesaw State College,
P.O. Box 444, Marietta, GA 30061

Editor:

Let $ABCDEF$ be a regular hexagon with sides of length n, a positive integer; it can be tiled with $6n^2$ equilateral triangles with sides of length 1. Now let the hexagon be tiled with $3n^2$ "calissons," i.e., rhombi formed by joining two of the triangles along a common side. David and Tomei [**1**] showed that such a tiling must contain exactly n^2 calissons with each of the three possible orientations. The purpose of this letter is to give a simpler proof of this fact.

Let a be the number of calissons with sides parallel to FA and AB, b the number with sides parallel to AB and BC, and c the number with sides parallel to BC and CD. Let $L_0, \ldots, L_{2n}$ be equally spaced parallel lines such that AB lies on L_0 and DE on L_{2n}. For $1 \le i \le 2n$, let k_i be the number of calissons with one side on L_{i-1} and another on L_i.

It is clear that $k_1 = n$ and $k_i = k_{i+1}$ for $1 \le i < 2n$. Hence $k_1 = k_2 = \cdots = k_{2n} = n$ and $a + b = k_1 + \cdots + k_{2n} = 2n^2$. By symmetry we also have $b + c = 2n^2$ and $c + a = 2n^2$, whence $a = b = c = n^2$.

The author was partly supported by NSF grant DMS 8802856.

REFERENCE

1. Guy David and Carlos Tomei, The problem of the calissons, this MONTHLY, 96 (1989) 429–431.

FRED GALVIN
Department of Mathematics,
University of Kansas,
Lawrence, KS 66045

Editor:

It was gratifying to read the article [**1**] in which George P. Richardson decried the omission of the midpoint rule in calculus textbooks. Indeed, an additional argument can be made for its use as a complement to the trapezoidal rule.

In estimating $\int_a^b f(x)\,dx$, let T_n and M_n be the trapezoidal and midpoint rule approximations to the integral respectively, where the interval $[a, b]$ is divided into n subintervals of equal length. In order to implement Romberg's method, it is necessary to compute a sequence of estimates T_n, doubling the value of n each time. If both the trapezoidal and midpoint rules are used, then one may take advantage of the following formula, which may be verified easily by the reader,

$$T_{2n} = (T_n + M_n)/2.$$

Consequently, once an initial value of the trapezoidal rule, say for $n = 1$, is established, function evaluations are done only when computing the midpoint rule. Since it is necessary to evaluate a function at a given point only once, the computation is extremely efficient.

REFERENCE

1. George P. Richardson, Reconsidering area approximations, this MONTHLY, 95 (1988) 754–757.

ROBERT O. STANTON
St. John's University
Jamaica, NY 11439

Editor:

Regarding Larry Cusick and Peter Tannenbaum's article, "Fixed Points of the Twisted Cyclic Shift Operator," *Amer. Math. Monthly* (Oct 89), the number of states of an $(n + 1)$-bit twisted cyclic shift counter is $2(n + 1)$ rather than 2 to the nth power as implied by the article.

These values agree only when $n = 3$ which is exactly the case in the example given in Fig. 4. of the article.

Unfortunately, the implied ability to reduce the complicated logic of an n-bit Gray code sequencer by replacing it with an $(n + 1)$-bit twisted cyclic shift operator is not valid.

WILLIAM J. BLEWETT,
Four Pi Systems,
10905 Technology Place,
San Diego, CA 92127

Editor:

I would like to point out that the results and the way to prove them, contained in the paper "A simple estimate of the error in linear approximation" by R.M. Gethner, this MONTHLY, Vol. 96, N° 6, pp. 522–523, are exactly the same as those presented on page 181 of the Classic "Introduction to Calculus and Analysis", by Courant and John, Vol. 1, 1965, Interscience.

Sincerely Yours,

ROBERTO A. MACÍAS
INTEC-PEMA
Guemes 3450
3000 Santa Fe
ARGENTINA

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

Open Problems in Grid Labeling

JOSEPH A. GALLIAN*
Department of Mathematics and Statistics, University of Minnesota, Duluth, MN 55812

Let P_m denote the path with m vertices and C_m the cycle with m vertices. From these we may form three kinds of grid graphs using the Cartesian product:

$P_m \times P_n$ a planar grid with m copies of P_n (rows) and n copies of P_m (columns)
$P_m \times C_n$ a grid on a cylinder with m copies of C_n (horizontally) and n copies of P_m (vertically)
$C_m \times C_n$ a grid on a torus with m copies of C_n (latitudinally) and n copies of C_m (longitudinally)

where, by the Cartesian product $G_1 \times G_2$, we mean the graph with vertex set $\{(x, y) | x \in G_1, y \in G_2\}$ and two vertices (x_1, y_1) and (x_2, y_2) of $G_1 \times G_2$ are adjacent if $x_1 = x_2$ and y_1 and y_2 are joined by an edge or if $y_1 = y_2$ and x_1 and x_2 are joined by an edge.

In the past few years a number of authors, including four undergraduates, have investigated the problem of whether these grids can be labeled in various ways. In particular, a connected graph with v vertices and e edges is called *graceful* if it is possible to label the vertices x with distinct integers $f(x)$ in $\{0, 1, 2, \ldots, e\}$ so that when each edge xy is labeled $|f(x) - f(y)|$, the resulting edge labels are distinct (and thus form the entire set $\{1, 2, 3, \ldots, e\}$). One type of graceful labeling is interesting enough to warrant its own name. Thus, a graceful labeling f is called an *α-valuation* if there is an integer k such that for any edge xy, either $f(x) \leqslant k < f(y)$ or $f(y) \leqslant k < f(x)$. The notions of graceful graphs and α-valuations were introduced by Rosa in 1967 [**17**] (although the term "graceful" was coined by Golomb [**10**] and popularized by Martin Gardner [**9**]). Problems about graceful graphs have appeared in this section of the MONTHLY on several occasions [**5**], [**11**], [**14**].

If f is an α-valuation on a graph G with vertex set V and k is the smaller of the labels $f(x)$ and $f(y)$ for the edge xy with $|f(x) - f(y)| = 1$ then G is bipartite with parts $\{x \in V | f(x) > k\}$ and $\{x \in V | f(x) \leqslant k\}$. A standard result in graph theory says that a bipartite graph cannot have an odd cycle. Thus, the graph

*This paper is based on work supported by the National Science Foundation under Grant No. 8709428.

$C_m \times P_n$ does not have an α-valuation when m is odd and the graph $C_m \times C_n$ does not have an α-valuation when m or n is odd.

In 1980, as an additive analog of graceful graphs, Graham and Sloane [**13**] defined a connected graph with e edges to be **harmonious** if it is possible to label the vertices x with distinct integers $f(x)$ in $\{0, 1, 2, \ldots, e-1\}$ so that when edge xy is labeled $f(x) + f(y)$, the resulting edge labels are distinct modulo e.

Grace [**12**] introduced what appears to be a stronger form of harmonious graph by calling a connected graph with e edges **sequential** if it is possible to label the vertices x with distinct integers $f(x)$ in $\{0, 1, 2, \ldots, e-1\}$ so that when edge xy is labeled $f(x) + f(y)$ the set of edge labels is a block of e consecutive integers. Obviously, these edge labels are distinct modulo e. Surprisingly, there is no known example of a harmonious graph which doesn't admit a sequential labeling.

Although there have been upwards of 150 papers written on graceful and harmonious graphs, there are few general results. Indeed, even for problems as narrowly focused as the ones involving the aforementioned grids, the labelings have been hard-won and involve a large number of cases. The progress to date in grid labeling is summarized in the table below. We invite readers to fill the gaps.

Surveys of other results and open problems on graph labeling are given in [**2**], [**7**], and [**8**]. Applications are discussed in [**3**] and [**4**].

In the table below we always exclude the trivial cases involving P_1 and C_1. The entry NO means the labeling is impossible; an asterisk refers to the earlier observation that graphs with α-valuations can not have odd cycles.

Graph	α-valuation	Graceful	Harmonious
$P_m \times P_n$	all **[16, 1, 15]**	all **[16, 1, 15]**	all except $(m, n) = (2, 2)$ **[15]**
$C_{2m} \times P_{2n}$	all **[15]**	all **[15]**	all except $(m, n) = (2, 1)$ **[15]**
$C_{4m} \times P_{2n+1}$	all **[15]**	all **[15]**	all **[15]**
$C_{4m+2} \times P_{2n+1}$	?	?	?
$C_{2m+1} \times P_n$	NO*	$n = 2$ **[6]**; $n > 2$?	all **[13]**
$C_{4m} \times C_{2n}$	all **[15]**	all **[15]**	$m = 1$, $n > 1$ **[15]**; otherwise?
$C_{4m} \times C_{2n+1}$	NO*	?	?
$C_{4m+2} \times C_{2n+1}$	NO*	?	NO [**13**, Theorem 11]
$C_{2m+1} \times C_{2n+1}$	NO*	NO **[16]**	NO [**13**, Theorem 11]

REFERENCES

1. B. D. Acharya and M. K. Gill, On the index of gracefulness of a graph and the gracefulness of two-dimensional square lattice graphs, *Indian J. Math.*, 23 (1981) 81–94.
2. G. S. Bloom, A chronology of the Ringel-Kotzig conjecture and the continuing quest to call all trees graceful, *Ann. N.Y. Acad. Sci.*, 326 (1979) 32–51.
3. G. S. Bloom and S. W. Golomb, Numbered complete graphs, unusual rulers, and assorted applications in Theory and Applications of Graphs, Lecture Notes in Math., 642, Springer-Verlag, New York (1978) 53–65.
4. _____, Applications of numbered undirected graphs, *Proc. IEEE*, 65 (1977) 562–570.
5. I. Cahit, Are all complete binary trees graceful, *Amer. Math. Monthly*, 83 (1976) 35–37.
6. R. Frucht and J. A. Gallian, Labeling prisms, *Ars Combinatoria*, 26 (1988) 69–82.
7. J. A. Gallian, Labeling prisms and prism related graphs, *Congressus Numerantium*, 59 (1987) 89–100.

8. J. A. Gallian, A survey—recent results, conjectures and open problems in labeling graphs, *J. Graph Theory*, to appear.
9. M. Gardner, Mathematical games: the graceful graphs of Solomon Golomb, or how to number a graph parsimoniously, *Scientific American* 226, 3 (1972) 108–112; 226, 4 (1972) 104; 226, 6, (1972) 118.
10. S. W. Golomb, How to number a graph, in Graph Theory and Computing, R. C. Read, ed., Academic Press, New York (1972) 23–37.
11. _____, The largest graceful subgraph of the complete graph, *Amer. Math. Monthly*, 81 (1974) 499–501.
12. T. Grace, On sequential labelings of graphs, *J. Graph Theory*, 7 (1983) 195–201.
13. R. L. Graham and N. J. A. Sloane, On additive bases and harmonious graphs, *SIAM J. Alg. Disc. Math.*, 1 (1980) 382–404.
14. R. Guy, Monthly research problems, 1969–1979, *Amer. Math. Monthly*, 86 (1979) 847–852.
15. D. Jungreis and M. Reid, Labeling grids, preprint.
16. A. Rosa, On cyclic decompositions of the complete graph (Slovak), CSc. dissertation, Bratislava, 1965.
17. _____, On certain valuations of the vertices of a graph, in Theory of Graphs (Internat. Symposium, Rome, July 1966), Gordon and Breach, N.Y. and Dunod Paris (1967) 349–355.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

A Theorem on the Numerators of the Bernoulli Numbers

KURT GIRSTMAIR
Institut für Mathematik, Universität Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria

The *Bernoulli numbers* B_m, $m = 0, 1, 2, \ldots$ are useful in several fields of mathematics, such as number theory, combinatorics, and numerical analysis. For odd $m \geq 3$, $B_m = 0$. For even $m \geq 2$, write $B_m = N_m/D_m$, where N_m, D_m are relatively prime integers, $D_m \geq 1$. The *denominators* D_m are well known, due to a theorem of *von Staudt* and *Clausen* of 1840. In fact, D_m is the product of all primes p with $p - 1 | m$. The *numerators* N_m are less well known. However, N_m deserves special interest; for instance, many fundamental results concerning *Fermat's Conjecture* rest on arithmetical properties of N_m (cf. [**2**]). One of the few *explicit* results on divisors of N_m is the following:

THEOREM. *Let $m \geq 2$ be even, and p a prime with $p - 1 \nmid m$. If p^r divides m ($r \geq 1$), then p^r divides N_m, too.*

This theorem also goes back to *von Staudt*, who published a proof in a hardly accessible "Universitätsschrift" ("De numeris Bernoullianis commentatio altera," Erlangen 1845). It was rediscovered by several 19th century mathematicians such as Sylvester [**7**], Adams [**1**], or Lipschitz [**5**]. Misattributions were made even in recent times (cf. [**6**]).

A number of proofs of this theorem are known. More "advanced" proofs are based on p-adic methods (e.g. [**4**]). "Elementary" proofs use subtle identities of power series, involving the generating function

$$f(t) = t/(e^t - 1) = \sum_{m=0}^{\infty} B_m t^m/m!$$

of the Bernoulli numbers, and congruence considerations (cf. [**2**], [**6**], [**3**]). However, it is not easy to grasp the ideas behind these proofs.

The aim of the present note is to convey a better understanding of the theorem within the framework of *algebraic* number theory, more precisely, of *cyclotomy*. Let $n \geq 2$ be an integer, and let $\zeta = e^{2\pi i/n}$ (i.e., a primitive nth root of unity). We consider the numbers B_m as special cases of more general "cyclotomic" Bernoulli numbers $B_{m,k}$, $k = 0, 1, \ldots, n - 1$, defined by means of the generating functions

$$f_k(t) = t/(\zeta^k \cdot e^t - 1) = \sum_{m=0}^{\infty} B_{m,k} t^m/m!.$$

Indeed, $f_0 = f$, hence $B_{m,0} = B_m$.

One source of our proof of the theorem is a *divisibility property* of $B_{m,k}$ for $k \neq 0$. It reads as follows.

I. *For each $m \geq 2$, $n \geq 2$, the numbers $(2n)^m B_{m,k}/m, k = 1, \ldots, n - 1$, are algebraic integers.*

Let us briefly recall the relevant facts about algebraic integers (cf. [3, pp. 66–68]).

(A) An *algebraic integer* is a complex number that is a root of a polynomial with coefficients in $\mathbb{Z}$ and leading coefficient = 1.

(B) The set of all algebraic integers is a *ring* (with the usual operations $+, -, \times$).

(C) All algebraic integers contained in $\mathbb{Q}$ are in $\mathbb{Z}$.

The second source of our proof is a relation between the numbers $B_{m,k}$ and B_m, namely,

$$\text{II.} \qquad \sum_{k=1}^{n-1} B_{m,k} = (n^m - 1)B_m \qquad (m \geq 0).$$

Indeed, I, II, and property (B) show that $(2n)^m(n^m - 1)B_m/m$ is an algebraic integer. Due to (C), this number is in $\mathbb{Z}$. Suppose now that $m \geq 2$ is even, $p^r | m$ ($r \geq 1$), and $p - 1 \nmid m$. In particular, $p \neq 2$. We can choose the number n as a *primitive root* modulo p (cf. [3, p. 41]). This means that $n \not\equiv 0 \bmod p$, and that $n^q \not\equiv 1 \bmod p$ whenever $p - 1 \nmid q$ ($q \geq 0$). In particular, p divides neither $2n$ nor $n^m - 1$, so p^r must divide N_m.

Proof of I. As a first step, we write the numbers $B_{m,k}$ in a "trigonometric" shape. For this purpose observe that

$$(e^{2it} + 1)/(e^{2it} - 1) = -i \cdot \cot(t) = 1 + 2/(e^{2it} - 1).$$

Since $f_k(t) = t/(e^{2i(\pi k/n - it/2)} - 1)$ we get

$$f_k(t) = (-it/2) \cdot \cot(\pi k/n - it/2) - t/2, \qquad k = 0, \ldots, n-1.$$

Now Taylor expansion of either side yields

$$B_{m,k} = m(-i/2)^m \cot^{(m-1)}(\pi k/n),$$

for $m \geq 2$ and $k = 1, \ldots, n-1$.

The occurrence of m on the right side is the decisive point in the proof of I. Actually we shall show that $n^m(-i)^m \cot^{(m-1)}(\pi k/n)$ is an algebraic integer for each $m \geq 1$. In the case $m = 1$, write $-i \cdot \cot(\pi k/n) = (\zeta^k + 1)/(\zeta^k - 1)$. Furthermore ζ^k is a root of the polynomial

$$(x^n - 1)/(x - 1) = x^{n-1} + \cdots + 1 = \prod_{l=1}^{n-1} (x - \zeta^l),$$

and thus an algebraic integer. On evaluating this polynomial at $x = 1$, we get

$$n = \prod_{l=1}^{n-1} (1 - \zeta^l).$$

Hence $-in \cdot \cot(\pi k/n) = -(\zeta^k + 1)\prod_{l \neq k}(1 - \zeta^l)$ *is an algebraic integer, by* (*B*).

When we differentiate the identity

$$-\cot^{(1)}(t) = 1 + \cot(t)^2 = 1 - (-i \cdot \cot(t))^2$$

repeatedly, we find that $(-i)^m \cot^{(m-1)}(t)$ is a polynomial in $-i \cdot \cot(t)$ of degree m with coefficients in $\mathbb{Z}$. For this reason $(-in)^m \cot^{(m-1)}(\pi k/n)$ is expressible as an integral polynomial in $-in \cdot \cot(\pi k/n)$, and (B) proves I.

Proof of II. The relation II between cyclotomic and ordinary Bernoulli numbers is the result of a corresponding relation between the generating functions. This, in turn, is in substance nothing but the decomposition of $1/(x^n - 1)$ into partial fractions, which has the shape

$$1/(x^n - 1) = \sum_{l=0}^{n-1} c_l/(x - \zeta^l), \text{ with } c_l \in \mathbb{C}.$$

There are standard methods to determine the coefficients c_l, but we shall do it in an *ad hoc* way here. First multiply the above equation by $x - \zeta^k$. Then put $x = \zeta^k$ and obtain

$$1\Big/\prod_{l\neq k}(\zeta^k - \zeta^l) = c_k = \zeta^k\Big/\prod_{l=1}^{n-1}(1 - \zeta^l).$$

As we have seen in the proof of I, the final denominator equals n, so $c_k = \zeta^k/n$. Hence

$$n/(x^n - 1) - 1/(x - 1) = \sum_{k=1}^{n-1} 1/(\zeta^k x - 1).$$

The substitution $x \mapsto e^t$ shows that

$$f_0(nt) - f_0(t) = \sum_{k=1}^{n-1} f_k(t),$$

and comparison of coefficients yields II.

REFERENCES

1. J. C. Adams, Table of the values of the first sixty-two numbers of Bernoulli, *J. reine angew. Math.*, 85 (1878) 269–272.
2. Z. I. Borevich and I. R. Shafarevich, Number Theory, Academic Press, New York, 1966.
3. K. Ireland and M. Rosen, A Classical Introduction to Modern Number Theory, Springer-Verlag, New York, Heidelberg, Berlin, 1982.
4. S. Lang, Cyclotomic Fields, Springer-Verlag, New York, Heidelberg, Berlin, 1978.
5. R. Lipschitz, Beiträge zu der Kenntniss der Bernoullischen Zahlen, *J. reine angew. Math.*, 96 (1884) 1–16.
6. H. Rademacher, Topics in Analytic Number Theory, Springer-Verlag, Berlin, Heidelberg, New York, 1973.
7. J. J. Sylvester, Sur une propriété de nombres premiers qui se rattache au théorème de Fermat, *C.R. Acad. Paris*, 52 (1861) 161–163.

Another Proof of the Quadratic Reciprocity Theorem?

RICHARD G. SWAN
Department of Mathematics, University of Chicago, Chicago, IL 60637

Here is a short proof of quadratic reciprocity based on an idea of Milgram [2]. Since this theorem probably holds the world's record for the number of known proofs, it is hard to believe that this one is really new, but I have not been able to

find this particular version in the literature. I would like to thank Milgram for pointing out the paper [1] which contains the theorem proved below. The proof in [1] makes use of Gauss' lemma rather than the norm calculation used here.

Let σ_p be a primitive pth root of 1 and set $K_p = \mathbb{R} \cap \mathbb{Q}(\sigma_p)$. The element $1 - \sigma_p$ is prime in $\mathbb{Q}(\sigma_p)$ with norm p. Let π_p be the norm of $1 - \sigma_p$ in K_p. Then $N(\pi_p) = p$ and $\pi_p = (1 - \sigma_p)(1 - \sigma_p^{-1}) = 2 - \sigma_p - \sigma_p^{-1}$. If p and q are primes consider $\eta = \pi_p - \pi_q$ in $L = K_pK_q$. It is a unit since $\eta = \sigma_q^{-1}(1 - \sigma_p\sigma_q)(1 - \sigma_p^{-1}\sigma_q)$. Since $\eta \equiv \pi_p \pmod{\pi_q}$, $N(\eta) \equiv N(\pi_p) \pmod{q}$. Now $N(\pi_p) = N_{K_p/Q}N_{L/K_p}(\pi_p) = N_{K_p/Q}(\pi_p^{(q-1)/2}) = p^{(q-1)/2} \equiv (p/q) \bmod q$ by Euler's criterion. But $N(\eta) = \pm 1$, so we have proved the following result.

THEOREM. *$N(\pi_p - \pi_q) = (p/q)$ for odd primes p and q.*

Interchanging p and q now gives quadratic reciprocity.

COROLLARY. *$(p/q)(q/p) = (-1)^{(p-1)(q-1)/4}$ for odd primes p and q.*

When $q = 2$, we can consider instead $\tau = 2 - \pi_p = \sigma_p + \sigma_p^{-1}$ in K_p. The same argument gives $N(\tau) = (2/p)$ but switching p and 2 is useless. Instead consider $f_n = x^n + \cdots + 1 + \cdots + x^{-n} = g_n(z)$ where $z = x + x^{-1}$. Then $zg_n(z) = g_{n-1}(z) + g_{n+1}(z)$ so $g_{n+1}(0) = -g_{n-1}(0)$ showing that $g_n(0) = 1$ for $n \equiv 0, 1 \bmod 4$ and $g_n(0) = -1$ otherwise. Since g_n is the minimal polynomial of τ for $n = (p - 1)/2$, we see that $(2/p) = N(\tau) = (-1)^n g_n(0) = (-1)^{(p^2-1)/8}$ as required.

REFERENCES

1. M. Gerstenhaber, The 152-nd proof of the law of quadratic reciprocity, this MONTHLY, 70 (1963) 397–398.
2. R. J. Milgram, Odd index subgroups of units in cyclotomic fields and applications, Springer Lecture Notes 854, 1980, pp. 269–298.

Adding up to Powers

EDWARD A. BENDER
Department of Mathematics, University of California—San Diego, La Jolla, CA 92093

FRED KOCHMAN
Institute for Defense Analyses, Princeton, NJ 08540

DOUGLAS B. WEST
Department of Mathematics, University of Illinois, Urbana, IL 61801

Here is a short combinatorial proof that a somewhat mysterious way of generating the kth powers of the natural numbers actually works. It generalizes the well-known fact that the sum of the first n odd numbers is n^2. The proof makes a nice presentation in an introductory combinatorics course. It uses recurrence relations, proof by 1-1 correspondence, and counting of lattice paths in a problem that students are likely to find appealing.

After discussing the kth powers, we will discuss a generalization of this procedure. We can still explain the resulting numbers as values of polynomials. We will see, however, that generating the kth powers is a very special case.

1. Generating the Perfect Powers. The procedure is as follows. Begin by writing down the natural numbers in sequence, calling this row 1. Next, strike out every kth term. Take partial sums of the surviving sequence to generate row 2. In general, generate row $j+1$ from row j by striking out every $k+1-j$th term and taking partial sums of the surviving sequence. The claim is that row k consists of precisely the kth powers of the natural numbers, in sequence. The procedure is illustrated below for $k = 4$. Note that the successive deletions make the numbers appear in wedges; a number is in wedge n if the first deletion to its right is the nth deletion in that row.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		3	6	11	17	24		33	43	54		67	81	96		113	131	150	
1		4		15	32			65	108			175	256			369	500		
1				16				81				256				625			

The numbers in the major diagonal of each wedge have a recognizable form. Let $a_{j,n}$ be the number in row j of the nth major diagonal. Then in fact $a_{j,n} = n^j\binom{k}{j}$. We prove this by showing that the $a_{j,n}$ and the numbers $b_{j,n} = n^j\binom{k}{j}$ satisfy the same recurrence. First we study the $b_{j,n}$.

LEMMA 1. *The values* $b_{j,n} = n^j\binom{k}{j}$ *satisfy the recurrence* $b_{j,n} = \sum_{i=1}^{j}\binom{k-i}{j-i}b_{i,n-1}$.

Proof. Consider the sequences formed from the integers $\{0,\ldots,n\}$. Let $S_{k,j}$ be the sequences that have k terms, of which j are nonzero. Both sides of the claimed equality count $S_{k,j}$. To show $|S_{k,j}| = n^j\binom{k}{j}$, simply choose the positions for the non-zero entries and fill them in. To show the sum also counts $S_{k,j}$, partition $S_{k,j}$ by the number of positions containing terms greater than 1. This number may be anywhere from 0 to j; if it is i, the positions of terms exceeding 1 can be chosen in $\binom{k}{i}$ ways, the positions can be filled in $(n-1)^i$ ways, and the $j-i$ 1's can be placed in $\binom{k-i}{j-i}$ ways. □

To prove the desired theorem, we need only verify the initial conditions and show that the $a_{j,n}$ satisfy the same recurrence. First, we need a row corresponding to $j = 0$. So, add a 0th row to the display, consisting entirely of 1's. In particular, $a_{0,n} = 1$. The wedge $n = 0$ is entirely 0 except $a_{0,0} = 1$.

THEOREM 1. $a_{j,n} = n^j\binom{k}{j}$.

Proof. The proof is by induction; we have already verified the basis. We need only show $a_{j,n} = \sum_{i=0}^{j}\binom{k-i}{j-i}a_{i,n-1}$ for $n \geq 1$. Note that every entry in the display is the sum of the entry to its left and the entry above it. If we imagine an $a-1$th row consisting of 0's, this means every $a_{j,n}$ can be broken down into contributions from entries in the main diagonal of the $n-1$st wedge. We need only determine how many times each $a_{i,n-1}$ contributes to each $a_{j,n}$. This is precisely the number of paths from $a_{i,n-1}$ to $a_{j,n}$ that always move rightward or downward at each step.

Due to the deletions in each row, there is no element below $a_{i,n-1}$, so every path from $a_{i,n-1}$ must take its first step to the right. From there, the paths to $a_{j,n}$ are precisely the lattice paths counted by the appropriate binomial coefficients. These paths enter the nth wedge in the ith diagonal, and they end at the kth diagonal, so the total number of steps is $k - i$. The number of steps downward is $j - i$, so the correct coefficient is indeed $\binom{k-i}{k-j}$. (In fact, more generally, the jth entry in the lth diagonal of the nth wedge is $\sum_{i=0}^{j}\binom{l-i}{j-i}n^j\binom{k}{j}$.) □

In particular, $a_{k,n} = n^k$, as desired.

2. A Generalization. The procedure generating the powers can be generalized to yield other polynomials. As before, begin with row 0 having a 0 in position 0 and 1's thereafter. At row j, choose a modulus m_j and a remainder r_j with $1 \leqslant r_j \leqslant m_j$. Create row $j + 1$ by 1) deleting from row j the entries $r_j + sm_j$ for $s \geqslant 0$, and 2) computing partial sums of what remains. We call this the *addition procedure*. Row 1 is always the natural numbers, so we don't bother to specify m_0, r_0. To generate the powers of the natural numbers, we used the addition procedure with $m_j = k + 1 - j$ and $r_j = m_j$. We claim that, for any choice of $\{m_j\}$ and $\{r_j\}$, the sequence arising at each row consists of the values of a finite number of polynomials, evaluated at consecutive integers. To prove this, we first review elementary facts about polynomials and binomial coefficients.

LEMMA 2. *Any polynomial $p(n)$ of degree j has a unique expression as a linear combination of $\binom{n}{0}, \ldots, \binom{n}{j}$.*

Proof. If the leading coefficient of $p(n)$ is c, let the coefficient of $\binom{n}{j}$ be $j!c$ and apply induction. □

LEMMA 3. *If a, b are constants and p is a polynomial of degree j with leading coefficient c, then $\sum_{s=0}^{t} p(as + b)$ is a polynomial in t of degree $j + 1$ with leading coefficient $ca^j/(j + 1)$.*

Proof. The expression $p(as + b)$ is a polynomial of degree j in s. Using the preceding lemma (and its proof), we express it as $\sum_{i=0}^{j} c_i\binom{s}{i}$, with $c_j = ca^j j!$. By interchanging the order of summation and applying the fundamental combinatorial identity $\sum_{s=0}^{t}\binom{s}{i} = \binom{t+1}{i+1}$, the desired sum becomes $\sum_{i=0}^{j} c_i\binom{t+1}{i+1}$. This is a polynomial in t of degree $j + 1$, with leading coefficient $c_j/(j + 1)! = ca^j/(j + 1)$. □

We say that a sequence $b_1, \ldots$ is a *polynomial sequence of degree k* if $b_i = p(i)$ for $1 \leqslant i \leqslant n$ for some polynomial p of degree k.

THEOREM 2. *Let $\{(r_j, m_j)\}$ be an instance of the addition procedure. For each j, there is a modulus M_j such that the congruence classes of positions mod M_j partition row j into polynomial subsequences of degree j. Furthermore, the leading coefficients of these polynomials are equal and depend only on $(m_1, \ldots, m_j)$.*

Proof. We prove this by induction on j; for $j = 1$ we may choose $M_1 = 1$ and the polynomial $p(i) = i$.

Consider $j \geqslant 1$, and suppose the claim holds for row j with polynomials $p_0, \ldots, p_{M_j - 1}$ having leading coefficients c. Let f_i denote the ith element of row $j + 1$. Let $a = \operatorname{lcm}(M_j, m_j)$, and define $M_{j+1} = a(1 - 1/m_j)$. Fix an integer r with

$0 \leqslant r < M_j$, and consider the subsequence $g_s = f_{sM_j+r}$. The difference $g_{s+1} - g_s$ is composed of contributions from the polynomials $\{p_l\}$. The choice of M_{j+1} accounts for the deletions from row j so that the pattern of contributions to $g_{s+1} - g_s$ from elements explained by a given p_l is the same for each value of s. Each polynomial p_l contributes a consecutive segment of its values, except that gaps arise due to deletions from row j. In particular, the contribution from p_l has the form $\sum_{i=1}^{d} p_l(n + \alpha_i)$, where $d \leqslant a/M_j$, the α_i's are d of the first a/M_j natural numbers, and the omissions depend only on r and l. For any value of n, this is the sum of d jth-degree polynomials in n. Hence it also has degree j, and it has leading coefficient dc.

The value of n used in this polynomial to compute the contribution of p_l to $g_{s+1} - g_s$ is $(a/M_j)s + \lfloor r/M_j \rfloor$. The contribution to $g_t - g_0$ from p_l is thus a partial sum of equally-spaced terms from a polynomial sequence of degree j. By the lemma, the sum is a polynomial in t of degree $j + 1$, with leading coefficient $dc(a/M_j)^j/(j + 1)$. Finally, we collect the contributions to g_t from all p_l and add the constant $g_0 = f_r$ to express g_s as a polynomial in s of degree $j + 1$. The leading coefficient is $c(a/M_j)^j/(j + 1)$ times the sum of the values d associated with the various p_l's. This sum is always M_{j+1}, because it equals the number of elements of row j added in to increase g_s to g_{s+1}. This completes the induction. □

With $M_1 = 1$ and $M_{j+1} = \operatorname{lcm}(M_j, m_j) \cdot (m_j - 1)/m_j$ as in the proof, we conclude that the leading coefficient for the polynomials explaining row j is $(1/j!)\prod_{i=1}^{j-1} M_{i+1}(\operatorname{lcm}(M_i, m_i)/M_i)^i$.

3. Additions and Differences. In the special case considered in Section 1, the polynomial $b_j(n) = \binom{k}{j} n^j$ is precisely the polynomial generated by the above procedure for row j when $m_j = r_j = k + 1 - j$ for $j = 1, \ldots, k$.

As another example, consider the addition procedure with $m_1 = 3$, $r_1 = 2$, $m_2 = 5$, $r_2 = 3$. The first three rows of the resulting display are

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17...
0	1		4	8		14	21		30	40		52	65		80	96	...
0	1		5			19	40		70	110			175		255	351	...

Taking the polynomial n for row 1, the two polynomials obtained for row 2 are $6\binom{n}{2} + 4n = 3n^2 + n$ and $6\binom{n}{2} + 7n + 1 = 3n^2 + 4n + 1$. In passing from row 2 to row 3, the pattern of deletions versus usage of polynomials repeats after 8 surviving terms of row 3, because $8 = \operatorname{lcm}(2, 5) \cdot (4/5)$. Thus eight polynomials suffice to explain the elements of row 3; the first two are $1200\binom{n}{3} + 1300\binom{n}{2} + 255n$ and $1200\binom{n}{3} + 1450\binom{n}{2} + 350n + 1$.

Note that we expressed these polynomials in terms of binomial coefficients. Once we know which values are covered by one of these polynomials, we can obtain the full polynomial quickly from $j + 1$ elements of row j by using the method of repeated differencing. Take the first $j + 1$ values of one congruence class mod M_j, the first of which will be the value of the desired polynomial at $n = 0$. Take the differences of consecutive values to get a sequence of j values. Continue this differencing to produce successively shorter sequences, the $j + 1$th sequence having only a single value. Let $c_0, \ldots, c_j$ denote the initial values in these $j + 1$ successive difference sequences. It is easily proved by induction that the tth difference

sequence consists of the first $j + 1 - t$ values of the polynomial $\Sigma_{i \geqslant t} c_i \binom{n}{i-t}$. In particular, the polynomial giving the initial sequence is $\Sigma c_i \binom{n}{i}$.

At this point, the curious reader will wonder what polynomials can arise from this procedure. In general, this appears to be difficult. However, suitable restrictions of the question yield interesting results.

First, note that winding up with a single polynomial as we did in Section 1 is a fairly special occurrence; it seems to require $M_k = 1$. Nevertheless, we can always guarantee it by choosing $m_j = k + 1 - j$. This still leaves a fair amount of variability in the choice of r_j, but since $0 \leqslant r_j < m_j$, this limits us to at most $k!$ polynomials for each value of k. Furthermore, the computation in the proof of Theorem 2 shows that these are all monic polynomials, which makes this class even more interesting.

Using the differencing procedure and the relationship between $\{\binom{n}{i}\}$ and $\{n^i\}$ via the Stirling numbers, it is a pleasant exercise to calculate some of these polynomials. For $k = 2$ the two polynomials are n^2 and $n^2 + n$. For $k = 3$, the six polynomials can be described concisely; n^3 has a different form from the other five, which are $n^3 + j(n^2 + n)/2 - n$, for $j = 1, 2, 3, 4, 5$. For $n = 4$, more formulas are necessary to explain all the polynomials, and the polynomials are not all distinct. We will close by listing those for which all coefficients are integers.

r_1	r_2	r_3	n^4	n^3	n^2	n
4	3	2	1			
1	3	2	1	1		
4	3	1	1	2	1	
3	2	1	1	3	3	1
2	1	2	1	2	1	
2	1	1	1	4	5	2

The appearance of the pure powers no longer looks quite so strange. Certainly there are some patterns here that are worth exploring.

Note added in proof: We have learned from C. T. Long and from this MONTHLY, 95 (1988) 708 that the process of our Section 1 is called Moessner's process and has been known since 1951. See Long's paper in this MONTHLY, 73 (1966) 846–851, for references and a proof of a certain generalization, different from our generalization.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

A One-Sentence Proof That Every Prime $p \equiv 1 \pmod 4$ Is a Sum of Two Squares

D. ZAGIER
Department of Mathematics, University of Maryland, College Park, MD 20742

The involution on the finite set $S = \{(x,y,z) \in \mathbb{N}^3 : x^2 + 4yz = p\}$ defined by

$$(x,y,z) \mapsto \begin{cases} (x + 2z,\ z,\ y - x - z) & \text{if } x < y - z \\ (2y - x,\ y,\ x - y + z) & \text{if } y - z < x < 2y \\ (x - 2y,\ x - y + z,\ y) & \text{if } x > 2y \end{cases}$$

has exactly one fixed point, so $|S|$ is odd and the involution defined by $(x,y,z) \mapsto (x,z,y)$ also has a fixed point. □

This proof is a simplification of one due to Heath-Brown [1] (inspired, in turn, by a proof given by Liouville). The verifications of the implicitly made assertions—that S is finite and that the map is well-defined and involutory (i.e., equal to its own inverse) and has exactly one fixed point—are immediate and have been left to the reader. Only the last requires that p be a prime of the form $4k + 1$, the fixed point then being $(1,1,k)$.

Note that the proof is not constructive: it does not give a method to actually find the representation of p as a sum of two squares. A similar phenomenon occurs with results in topology and analysis that are proved using fixed-point theorems. Indeed, the basic principle we used: "The cardinalities of a finite set and of its fixed-point set under any involution have the same parity," is a combinatorial analogue and special case of the corresponding topological result: "The Euler characteristics of a topological space and of its fixed-point set under any continuous involution have the same parity."

For a discussion of constructive proofs of the two-squares theorem, see the Editor's Corner elsewhere in this issue.

REFERENCE

1. D. R. Heath-Brown, Fermat's two-squares theorem, *Invariant* (1984) 3–5.

Inverse Functions and their Derivatives

ERNST SNAPPER
Department of Mathematics and Computer Science, Dartmouth College, Hanover, NH 03755

If the concept of inverse function is introduced correctly, the usual rule for its derivative is visually so obvious, it barely needs a proof. The reason why the standard, somewhat tedious proofs are given is that the inverse of a function $f(x)$ is

usually graphed by flipping the graph of $f(x)$ over the 45° line. And although this procedure is formally correct, it hides the visual obviousness of the formula for the derivative of the inverse. In short, graphing the inverse by flipping over the 45° line is an error in pedagogy.

The concept of the inverse of a one-to-one and onto function should first be explained on the set-theoretic level. This means that domains (of definition) and ranges should not be restricted to subsets of the real numbers, but arbitrary sets, and also subsets of the plane and 3-space should be used. It should be stressed that the inverse is obtained by simply interchanging domain and range. If this is done properly, the students should feel irritated that their time is being wasted on such trivialities.

What does this mean for the graph of the inverse of a real variable function $f(x)$ in one variable? Suppose the graph of $f(x)$ is given:

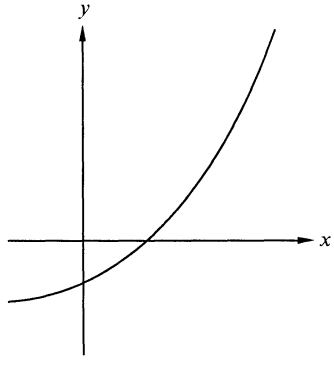


FIG. 1. The graph of $y = f(x)$.

How does one find the graph of the inverse $g(x)$ of $f(x)$? Since the students know that this is a matter of interchanging domain and range, they will quickly and correctly say: Interchange the x-axis and y-axis. Hence the graph is:

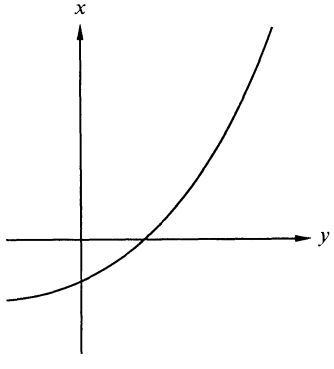


FIG. 2. The graph of the inverse $y = g(x)$.

The instructor should stress that the fact that the two axes are now in an unnatural position is entirely immaterial. This is the graph of the inverse which is helpful and effective. For example, the standard theorem that $f(x)$ is strictly increasing (decreasing) iff its inverse has the same property is now visually obvious. The same holds for all other theorems. Let us work it out for the derivative.

The students have learned that the derivative of $f(x)$ at a point P on the curve $y = f(x)$ is $\tan \alpha$, where α is the angle which the positive x-axis makes with the upward direction of the tangent of $y = f(x)$ at P:

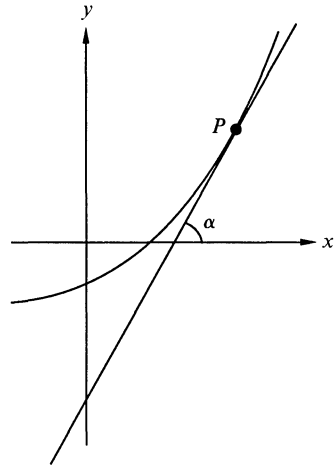


FIG. 3. $f'(P) = \tan \alpha$.

Interchanging the x-axis and y-axis gives, by the same rule, that the derivative of the inverse $g(x)$ of $f(x)$ at P is $\tan \beta$:

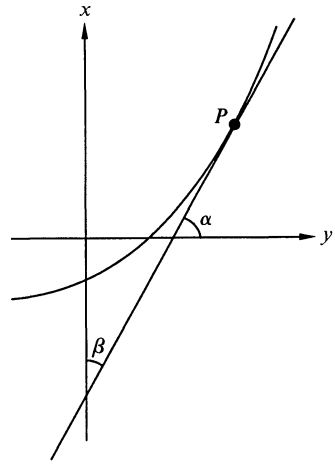


FIG. 4. $g'(P) = \tan \beta$.

Since $\alpha + \beta = 90°$, $\tan\beta = 1/\tan\alpha$ and the formula for the derivative of the inverse has been obtained.

This method of interchanging the x-axis and y-axis is also very efficient for viewing the graph of the inverse function with the axes in their *natural* position. Simply graph the function $f(x)$ on a piece of paper, putting a little pressure on the pencil, interchange the x and y and now look at the graph through the back of the paper with the axes in their natural position. What you see is the customary graph of the inverse of $f(x)$. It is very effective to do this in class, choosing for $f(x)$ some trigonometric function or a polynomial function such as $y = x^2$ or $y = x^3$. You can easily look through the back of the paper by holding it up to a window on a sunny day or up to a light. No visual aids are necessary or even helpful.

Of course, once the above has been done, there is nothing against making the students observe that the graph of the inverse of $f(x)$, viewed through the back of the paper, is the same as the graph of $f(x)$ flipped over the 45° line. This flipping over the 45° line can again be done easily with pencil and paper but, at least with some programs, computers can do it too. The fact remains, however, that for the purpose of understanding the laws which govern the inverse, nothing beats interchanging the x-axis and y-axis in the graph of $f(x)$ and doing nothing else.

Extending the Principal Axis Theorem to Fields Other Than $\mathbb{R}$

STEPHEN H. FRIEDBERG
Department of Mathematics, Illinois State University, Normal, IL 61761

Perhaps one of the most important theorems in all of linear algebra is the *principal axis theorem*, which states that a real symmetric matrix is orthogonally equivalent to a diagonal matrix; see, for example, [**1**, p. 337]. This theorem finds many applications in geometry and numerical analysis as well as statistics and physics, particularly in cases where symmetric bilinear forms arise. It is seldom pointed out that this theorem is not true, however, if no constraints are made on the underlying field $\mathbb{F}$. The first four examples illustrate the difficulty in extending the principal axis theorem to fields other than the field $\mathbb{R}$ of real numbers.

Example 1. Let $\mathbb{F} = \mathbb{C}$, the field of complex numbers, and

$$A = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}.$$

The characteristic polynomial of A is $f_A(x) = x^2$; so the only eigenvalue of A is 0. If A were diagonalizable, then A would equal the zero matrix.

Example 2. Let $\mathbb{F} = \mathbb{Q}$, the field of rational numbers, and

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}.$$

The characteristic polynomial of A is $f_A(x) = x^2 - 4x - 1$, which has no rational zeros. So A is not diagonalizable.

Example 3. Let $\mathbb{F} = \mathbb{Z}_2$, the field of two elements, 0 and 1. If

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix},$$

then $f_A(x) = x^2 + x + 1$. Because $f(0) = 1$ and $f(1) = 1$, we see that A has no eigenvalues, and hence A is not diagonalizable.

Example 4. If $\mathbb{F} = \mathbb{F}(\mathbb{C})$ is the field of algebraic numbers (the complex zeros of polynomials in $\mathbb{Q}[x]$), the matrix of Example 1 shows that the principal axis theorem fails.

Example 5. Let $\mathbb{F} = \mathbb{F}(\mathbb{R})$ be the subfield of $\mathbb{F}(\mathbb{C})$ consisting of the real algebraic numbers. Suppose that A is a symmetric matrix over $\mathbb{F}(\mathbb{R})$ and f_A is the characteristic polynomial of A. By the principal axis theorem, f_A factors over $\mathbb{R}$ in such a way that A is orthogonally diagonalizable. But because $\mathbb{F}(\mathbb{C})$ is algebraically closed, the eigenvalues must be elements of $\mathbb{F}(\mathbb{R})$. So A is orthogonally diagonalizable over $\mathbb{F}(\mathbb{R})$.

The preceding example shows that the principal axis theorem is valid when the underlying field is $\mathbb{F}(\mathbb{R})$. The main result of this article demonstrates that the principal axis theorem does not extend to any *finite* field. In fact, the situation there is hopeless in that there always exists a 2×2 symmetric matrix without *any* eigenvalues. Although purely algebraic proofs may be constructed, they assume results from field theory. The proof given here uses only a simple counting argument, and therefore can be given to students in an introductory linear algebra course.

THEOREM. *Let $\mathbb{F}$ be a finite field. Then there exists a 2×2 symmetric matrix (over $\mathbb{F}$) that possesses no eigenvalues.*

Proof. Suppose that $\mathbb{F}$ has n elements, and that A is an arbitrary 2×2 symmetric matrix, say,

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

The characteristic polynomial of A is

$$f_A(x) = x^2 - (a + c)x + (ac - b^2). \tag{1}$$

A quadratic monic polynomial $f(x)$ is reducible if and only if it has the form

$$f(x) = (x - d)(x - e).$$

If we first count the number of distinct choices of d and e, and then add the number of choices where $d = e$, it follows immediately that the number of reducible quadratic monic polynomials is

$$\binom{n}{2} + n = n\left(\frac{n+1}{2}\right). \tag{2}$$

We shall show that the number s of polynomials of the form in (1) is greater than the number of reducible quadratic monic polynomials. It will then follow that there exists a symmetric matrix A whose characteristic polynomial is irreducible, and, hence, A has no eigenvalues.

For each $k \in \mathbb{F}$, let $s(k)$ equal the number of ways of choosing the constant term $ac - b^2$ in (1), given that $a + c = k$. Notice that if $a + c = k$, then $ac - b^2 = ak - a^2 - b^2$. So $s(k)$ equals the number of elements in the set

$$\{ak - a^2 - b^2 : a, b \in \mathbb{F}\},$$

and

$$s = \sum_{k \in \mathbb{F}} s(k).$$

If we can show that $s(k) \geq (n + 1)/2$ for every k, with strict inequality for at least one k, then we will have that $s > n(n + 1)/2$, which by (2) equals the number of reducible quadratic monic polynomials. Note that for any k, if we let $a = 0$, then we obtain the constant term $-b^2$. Because $-b_1^2 = -b_2^2$ if and only if $b_1 = \pm b_2$, there are at least $1 + (n - 1)/2 = (n + 1)/2$ (distinct) negative squares. So $s(k) \geq (n + 1)/2$ for every k.

If every element of $\mathbb{F}$ is a (negative) square (e.g., if $\mathbb{F}$ has characteristic 2), then there are n such constant terms, and we are done. So suppose that there exists $e \in \mathbb{F}$ such that e is not a negative square. Let $k_0 = e + 1$, $a = 1$, and $b = 0$. Then

$$ak_0 - a^2 - b^2 = (e + 1) - 1 - 0^2 = e.$$

Thus

$$s(k_0) \geq (n + 1)/2 + 1 > (n + 1)/2,$$

and the proof is complete.

Problem. Classify exactly those fields for which the principal axis theorem is true.

REFERENCE

1. S. Friedberg, A. Insel, and L. Spence, Linear Algebra, Prentice-Hall, Englewood Cliffs, N.J., 1989.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, FREDERICK W. LUTTMAN, MARVIN MARCUS, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.

An asterisk () after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before July 31, 1990 at the MONTHLY PROBLEMS address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3367. *Proposed by John Kieffer, University of Minnesota, Minneapolis.*

Let A_0 and A_1 be disjoint infinite sets of positive integers whose union is the set Z^+ of all positive integers. For each infinite subset S of Z^+ let $f(S)$ be the set obtained from S by removing the least element s of S and then adjoining all the elements of A_{1-i} less than s if s is in A_i. Let $f^2(S) = f(f(S))$, $f^3(S) = f(f^2(S))$, etc. Given any positive integer k, prove that for sufficiently large n either

$$k \in f^n(A_0) \cap f^n(A_1) \quad \text{or} \quad k \notin f^n(A_0) \cup f^n(A_1).$$

E 3368. *Proposed by I. E. Leonard, University of Alberta, Edmonton, Alberta, Canada.*

Suppose $\alpha > 0$. Let us define a sequence $a_1, a_2, \ldots$ by taking a_1 as an arbitrary positive number and putting $a_{n+1} = a_n \exp\{-a_n^\alpha\}$ for $n = 1,2,\ldots$. Find the values of α for which the following series converge:

$$\text{(i)} \sum_{n=1}^{\infty} a_n, \quad \text{(ii)} \sum_{n=1}^{\infty} a_n^\alpha, \quad \text{(iii)} \sum_{n=1}^{\infty} a_n^{1-\alpha}.$$

E 3369. *Proposed by Jiro Fukuta, Aichi University, Aichi, Japan.*

Suppose we are given a piece of paper in the shape of an equilateral triangle ABC. Suppose P is a point in the intersection of the three open circular discs with diameters AB, BC, CA. If we fold the three corners of the paper in such a way that

the vertices coincide with P, we get a hexagon three sides of which are the creases formed and the other three sides of which are portions of the sides of triangle ABC.

Prove that the area and perimeter of this hexagon are both maximized when P is the centroid of ABC.

E 3370. *Proposed by Gavin Brown, University of New South Wales, Australia and Larry A. Shepp, A.T.& T. Bell Laboratories, Murray Hill, NJ.*

Suppose f and g are nonnegative continuous functions compactly supported on the real line. Hölder's inequality gives

$$\sup_x \int f(x-y)g(y)\,dy \le \|f\|_p \|g\|_q$$

for positive numbers p,q with $p^{-1} + q^{-1} = 1$. Prove or disprove that

$$\int \sup_y \{f(x-y)g(y)\}\,dx \ge \|f\|_p \|g\|_q.$$

E 3371. *Proposed by A. M. Garsia and B. A. Sethuraman, University of California at San Diego.*

Let S_n denote the symmetric group on the n symbols $1, 2, \ldots, n$. Given σ in S_n, let

$$M(\sigma) = \{i: 1 \le i \le n,\ i \ge \sigma(i)\}.$$

Put

$$P(\sigma) = |M(\sigma)|^{-1} \sum_{i \in M(\sigma)} \{i + \sigma(i)\}.$$

Prove that

$$\sum_{\sigma \in S_n} P(\sigma) = (n+1)!$$

E 3372. *Proposed by W. A. Bassali, Kuwait University, Safat.*

Suppose $\delta_n = \binom{2n}{n} 4^{-n}$ for $n = 0, 1, 2, \ldots$. Prove that

$$\int_0^{\pi/2} \sin^{-1}(\sin^2\theta)\,d\theta = \int_0^{\pi/2} \cos^{-1}\sqrt{\cos\theta}\,d\theta$$

$$= \int_0^1 \frac{\sinh^{-1} x}{x}\,dx = 2\int_0^{\pi/2} \frac{\theta\,d\theta}{\sqrt{\sec^2\theta + 1}}$$

$$= \frac{1}{2}\int_0^{\pi/2} \theta\sqrt{\operatorname{cosec}\theta + 1}\,d\theta = \sum_{n=0}^{\infty} \frac{(-1)^n \delta_n}{(2n+1)^2}$$

$$= \frac{\pi^2}{8} - \frac{1}{2}\sum_{n=0}^{\infty} \frac{\delta_{2n+1}}{(2n+1)^2} = \frac{\pi}{2}\sum_{n=0}^{\infty} \frac{\delta_n \delta_{2n+1}}{2n+1}.$$

For general $a_1, \ldots, a_n, b_1, \ldots, b_n$ we have $\Sigma(a_i - b_i)^2 \geq 0$, which may be rewritten

$$2 - 2\sum_{i=1}^{n} a_i b_i \geq 1 - \sum_{i=1}^{n} a_i^2 + 1 - \sum_{i=1}^{n} b_i^2. \tag{5}$$

We now assume (4). In view of (2) and (3), we see that (5) becomes

$$2\sum{}^* a_i a_j \geq \sum{}^* a_i a_j + \sum{}^* b_i b_j.$$

Dividing by 2 and applying the arithmetic-geometric mean inequality, we get the inequality of the problem. Furthermore, equality holds only when $b_i = \lambda a_i$ $(i = 1, \ldots, n)$ for some constant λ.

Editorial comment. Allen Pedersen and Walther Janous independently remarked that the result of the problem is a special case of Aczél's inequality. It can be found in D. S. Mitrinović, *Analytic Inequalities* (Springer-Verlag, 1970), p. 57, where there is also the generalization $(\Sigma^* a_i b_j)^p \geq [(\Sigma a_i)^p - \Sigma a_i^p][(\Sigma b_i)^p - \Sigma b_i^p]$, where $p > 1$ and the numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$ are nonnegative.

Also solved by J. Fukuta (Japan), W. Janous (Austria), F. Kubo and K. Kubo (Japan), L. Kuntz (student, West Germany), H. E. Lomeli (student, Mexico), O. P. Lossers (The Netherlands), W. Noll, A. Pedersen (Denmark), S. Philipp, G. S. Rogers, H.-J. Seiffert (West Germany), the Chico Problem Group, the University of South Alabama Problem Group, and the proposer.

Derivatives of Chebyshev Polynomials

E 3224 [1987, 681]. *Proposed by Bruce Reznick, University of Illinois at Urbana-Champaign, and John Todd, California Institute of Technology.*

Let $n \geq 3$ be an integer. Determine all real polynomials p of degree n having n real zeros $a_1 < a_2 < \cdots < a_n$ such that

$$\int_{a_k}^{a_{k+1}} |p(t)|\, dt \qquad (1 \leq k \leq n-1)$$

is independent of k.

Solution by Jim Delany and Kempton Huehn, California Polytechnic State University, San Luis Obispo. The first derivatives of the Chebyshev polynomials have this property. Except for scale changes and translations along the x-axis, these are the only such functions.

We first demonstrate that after normalization there is at most one such polynomial. Let $p(x)$ be such a function. The $n-1$ arches created by p and the x-axis enclose the same area. Since p has no multiple roots, the arches are alternately above and below the x-axis. We normalize p by translating to obtain $a_1 = 0$, scaling the x-axis to obtain $a_n = 1$, and scaling the y-axis so the area of each arch is 1 and p is positive between a_1 and a_2. This replaces $p(x)$ by

$$\frac{(a_n - a_1)\, p(a_1 + (a_n - a_1)x)}{\int_{a_1}^{a_2} p(t)\, dt}.$$

Suppose p and q are two such normalized polynomials, with zeros $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, respectively. Let $P(x) = \int_0^x p(t)\, dt$ and $Q(x) = \int_0^x q(t)\, dt$. By construc-

tion, $P(a_i) = Q(b_i) = 0$ if i is odd, and $P(a_i) = Q(b_i) = 1$ if i is even. Now, P and Q are polynomials of degree $n + 1$. We will show that $R = P - Q$ has more than $n + 1$ zeros, thus implying that $P - Q$ is identically zero and $p = q$ (since $p(0) = q(0) = 0$).

Since P, Q, P', Q' are all 0 at 0, R has a double zero at 0; i.e. x^2 divides R. Similarly, $P(1) = Q(1)$ and $P'(1) = Q'(1) = 0$ imply that $(x - 1)^2$ divides R. In addition to these four zeros of R, we need to guarantee $n - 2$ more.

We may assume $b_2 \le a_2$. Since $Q(b_2) = 1 \ge P(b_2)$ and $P(a_2) = 1 \ge Q(a_2)$, we have $R(a_2) \ge 0 \ge R(b_2)$. Thus R has a zero in $[b_2, a_2]$. In each of the intervals $[a_2, a_3], \ldots, [a_{n-2}, a_{n-1}]$, P ranges between the extremes 0 and 1, and Q is bounded by 0 and 1, so there is a point where they are equal. Counting $[b_2, a_2]$, we have zeros in $n - 2$ interior intervals. These are distinct unless a_k is the zero for the two intervals starting and ending at a_k. However, if R has a zero at a_k, then $Q(a_k) = P(a_k) = \{1 + (-1)^k\}/2$; thus a_k is one of the b_i and so $Q'(a_k) = P'(a_k) = 0$. Hence, if R has a zero at a_k, it is a double zero and we can count it once in each of the two intervals starting and ending at a_k. Summing up, we see that R has at least $n + 2$ zeros and hence is identically zero.

For existence, consider the Chebyshev polynomial of degree n, given by

$$T_n(x) = \tfrac{1}{2}\left[\left(x + \sqrt{x^2 - 1}\right)^n + \left(x - \sqrt{x^2 - 1}\right)^n\right]$$

$$= \begin{cases} \cos(n \cos^{-1} x) & \text{if } |x| \le 1 \\ \cosh(n \cosh^{-1} x) & \text{if } x > 1 \end{cases}$$

The first few such polynomials are 1, x, $2x^2 - 1$, $4x^3 - 3x$, $8x^4 - 8x^2 + 1$, and $16x^5 - 20x^3 + 5x$. The only relative extrema are $T_n(x) = \pm 1$, when $x = \cos(k\pi/n)$, for $k = 1, \ldots, n - 1$. When $k = 0$ or $k = n$, we also have $T_n(x) = \pm 1$, but the derivative is not 0. The derivative of $T_{n+1}(x)$ is a polynomial of degree n having the desired property. Its roots are $a_k = \cos[\pi(n + 1 - k)/(n + 1)]$ for $k = 1, \ldots, n$, where we have taken care to specify the zeros in increasing order.

Editorial comment. Shorter solutions were offered by O. P. Lossers and W. W. Meyer, who applied the following known result from approximation theory to $P = \int p$: the polynomial of fixed degree which best approximates a constant on an interval (in the uniform norm) is, up to translation and scaling, a Chebyshev polynomial. See, for example, E. W. Cheney, *Introduction to Approximation Theory*, McGraw Hill, 1966, p. 61.

Solved also by S. M. Gagola, Jr., O. P. Lossers (The Netherlands), W. W. Meyer, S. Philipp, L. Verde-Star (Mexico), and the proposers. One partial solution was received.

Points on a "Polynomial Approximation" to a Circle

E 3225 [1987, 786]. *Proposed by Bjorn Poonen, Winchester, Massachusetts.*

Let r be a fixed real number greater than 1, and let $s_n(x)$ and $c_n(x)$ denote the nth Maclaurin polynomials for the sine and cosine functions, respectively. If x_n is the smallest positive solution of

$$s_n(x)^2 + c_n(x)^2 = r,$$

show that

$$\lim_{n\to\infty} \frac{x_n}{n} = \frac{1}{e}.$$

Solution by Stan Philipp, Pennsylvania State University, Altoona. We shall show that if n is sufficiently large (depending on r), then

$$x_n > e^{-1}n \tag{1}$$

and

$$x_n < e^{-1}n^{1+1/n} = e^{-1}ne^{\log n/n} < e^{-1}n + \log n. \tag{2}$$

The method actually shows that $x_n = e^{-1}n + (\log n)/(2e) + O(1)$.

For real x put

$$T_n(x) = e^{ix} - \{c_n(x) + is_n(x)\} = \sum_{k=n+1}^{\infty} (ix)^k/k! \tag{3}$$

and

$$P_n(x) = c_n(x)^2 + s_n(x)^2 - r = |e^{ix} - T_n(x)|^2 - r. \tag{4}$$

To prove (1) note that for $0 \le x \le n/e$ we have

$$\begin{aligned} |T_n(x)| &\le \frac{x^n}{n!}\left\{\frac{x}{n+1} + \frac{x^2}{(n+1)(n+2)} + \cdots\right\} \\ &\le \frac{x^n}{n!}\left\{\frac{1}{e} + \frac{1}{e^2} + \cdots\right\} \\ &\le \frac{x^n}{n!} \le \frac{n^n}{e^n n!} = O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Therefore, if n is sufficiently large, we have $|T_n(x)| < \sqrt{r} - 1$ for $0 \le x \le n/e$ and so $P_n(x) < 0$ for $0 \le x \le n/e$ by (4). This proves (1).

To prove (2) note that if $x = e^{-1}n^{1+1/n}$, then by (3)

$$\begin{aligned} |T_n(x)| &\ge \frac{x^{n+1}}{(n+1)!}\operatorname{Re}\left\{1 + \frac{ix}{n+2} + \frac{(ix)^2}{(n+2)(n+3)} + \cdots\right\} \\ &= \frac{x^{n+1}}{(n+1)!}\left\{1 - \frac{x^2}{(n+2)(n+3)}\right. \\ &\qquad \left. + \frac{x^4}{(n+2)(n+3)(n+4)(n+5)} - \cdots\right\} \\ &\ge \frac{x^{n+1}}{(n+1)!}\left\{1 - \frac{x^2}{(n+2)(n+3)}\right\}, \end{aligned}$$

the last inequality being true because the terms in the preceding alternating series decrease in absolute value. (We have $x = e^{-1}n^{1+1/n} < n/2$ for $n \ge 6$.) Hence

$$|T_n(x)| \ge \frac{x^n}{n!}\frac{x}{n+1}\left(1 - \frac{1}{4}\right) \ge \frac{x^n}{n!}\frac{1}{e}\frac{3}{4} = \frac{n^n n}{e^n n!}\frac{3}{4e} > c\sqrt{n}$$

for a positive constant c, by Stirling's formula. Hence, if n is sufficiently large, $|T_n(e^{-1}n^{1+1/n})| > \sqrt{r} + 1$ and hence $P_n(e^{-1}n^{1+1/n}) > 0$ by (4). Thus (2) is proved.

Editorial comment. Several solvers verified that positive roots exist for each positive integer n by noting that $s_n(0)^2 + c_n(0)^2 = 1$ while

$$\lim_{x\to\infty} \left\{ s_n(x)^2 + c_n(x)^2 \right\} = \infty.$$

Solved also by P. Alsholm (Denmark), P. Bracken (Canada), O. P. Lossers (The Netherlands), T. S. Norfolk, the University of South Alabama Problem Group, and the proposer.

Parallel Ratios in a Semicircle

E 3259 [1988, 350]. *Proposed by Jordi Dou, Barcelona, Spain.* Let R be a semicircular region bounded by a line L and a semicircle S with center on L. Suppose P_1 and P_2 are given points in the interior of R. We wish to find parallel lines l_1, l_2 through P_1, P_2, respectively, such that

$$\frac{P_1C_1}{P_1D_1} = \frac{P_2C_2}{P_2D_2}, \tag{*}$$

where C_1, D_1 are the intersections of l_1 with S and L and C_2, D_2 are the intersections of l_2 with S and L. Give a necessary and sufficient condition on P_1, P_2 for such parallel lines to exist.

Solution by Emil Raychev, Portland, Maine. The necessary and sufficient condition is that the line P_1P_2 intersect L outside R or be parallel to L.

For necessity, suppose l_1 and l_2 are lines with the desired property. By the relation (*), the lines C_1C_2 and P_1P_2 pass through a common point M external to R on the line L. Hence the existence of l_1, l_2 implies that the line P_1P_2 intersects L outside R or is parallel to L.

Furthermore, if P_1P_2 meets L at M outside R (with P_2 between P_1 and M), let t be a line through M and tangent to S (at a point T). Then $MC_1 \cdot MC_2 = MT^2$ and $(MC_2/MC_1) = (MP_2/MP_1)$. Multiplying these yields

$$(MC_2)^2 = (MT)^2 \frac{MP_2}{MP_1}.$$

Therefore, it is sufficient to construct a point A on segment MT such that $MA = MT \cdot (MP_2/MP_1)$ and then to construct a segment MC_2 with C_2 on S such that $(MC_2)^2 = MT \cdot MA$.

To produce A, construct a line through P_2 parallel to P_1T; let A be the intersection of this with MT. The parallel lines cut MP_1 and MT in the same ratio, so MA has the desired length. To obtain C_2, construct a semicircle with diameter MT and erect the segment through A perpendicular to MT. This meets the semicircle at a point B with $(MB)^2 = MT \cdot MA$. Thus MB has the desired length MC_2, and the circle centered at M that has this radius meets S at the desired point C_2. C_1 is the other intersection of MC_2 with S.

If P_1P_2 is parallel to L, then the solution is obtained by constructing a chord C_1C_2 parallel to and having the same length as P_1P_2.

Solved also by C. W. Dodge, J. Fukuta (Japan), H. Guggenheimer, W. Janous (Austria), L. Kuipers (Switzerland), and the proposer.

Asymptotic Estimates for Convergents of a Continued Fraction

E 3264 [1988, 352]. *Proposed by Stanley Rabinowitz, Alliant Computer Systems Corp., Littleton, MA.*

Let P_n/Q_n be the nth convergent for the continued fraction

$$a = \frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{3+}\,\frac{1}{4+}\cdots,$$

i.e., let $P_1 = 1$, $Q_1 = 1$, $P_2 = 2$, $Q_2 = 3$, and

$$P_n = nP_{n-1} + P_{n-2}, \qquad Q_n = nQ_{n-1} + Q_{n-2} \qquad (n \geq 3).$$

(Cf. Hardy and Wright, *An Introduction to the Theory of Numbers*, Chapter 10.)

Give asymptotic estimates for P_n and Q_n.

Composite solution I by G. A. Edgar, Ohio State University, and Klaus Zacharias, Berlin. As n tends to ∞, P_n and Q_n are asymptotic to $I_1(2)n!$ and $I_0(2)n!$, respectively, where I_0, I_1 are the modified Bessel functions usually so designated. The numerical values are $I_1(2) = 1.59063685\ldots$ and $I_0(2) = 2.27958530\ldots$.

It should first be noted that the given continued fraction converges (cf. W. B. Jones and W. J. Thron, *Continued Fractions: Analytic Theory and Applications*, 1980, p. 88). Next, both P_n and Q_n satisfy the recurrence

$$A_n = nA_{n-1} + A_{n-2}. \tag{1}$$

If K_v and I_v denote the usual modified Bessel functions, then (see G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd edition, 1966, pp. 79–80)

$$K_{n+1}(x) = \frac{2n}{x}K_n(x) + K_{n-1}(x), \tag{2}$$

$$I_{n+1}(x) = \frac{-2n}{x}I_n(x) + I_{n-1}(x), \tag{3}$$

$$K_{n+1}(x)I_n(x) + K_n(x)I_{n+1}(x) = \frac{1}{x}. \tag{4}$$

From (2) and (3), we see that $A_n = K_{n+1}(2)$ and $A_n = (-1)^{n+1}I_{n+1}(2)$ are solutions of (1). Using (2)–(4) to check the initial conditions, we find that, for $n \geq 1$,

$$P_n = 2(-1)^{n+1}K_1(2)I_{n+1}(2) + 2I_1(2)K_{n+1}(2)$$
$$Q_n = 2(-1)^{n}K_0(2)I_{n+1}(2) + 2I_0(2)K_{n+1}(2).$$

From the definition of $I_n(2)$ [ibid, p. 77], we see that $I_n(2)$ tends to 0 as n tends to ∞. Furthermore [ibid, p. 80], $K_{n+1}(2) \sim \frac{1}{2}n!$ as n tends to ∞. This yields the asymptotics claimed for P_n, Q_n.

Solution II by Robert B. Israel, University of British Columbia. For any fixed k we show that P_n and Q_n have the asymptotic expansions

$$P_n \sim p \sum_{j=0}^{k} (-1)^j (n-j)!/j! = pn!\{1 - n^{-1} + [2n(n-1)]^{-1} - \cdots\},$$

$$Q_n \sim q \sum_{j=0}^{k} (-1)^j (n-j)!/j!,$$

where p and q are constants. More precisely, we show that if $n \geq k$, then $p\Sigma_{j=0}^{k}(-1)^j(n-j)!/j!$ is greater than P_n if k is even and less than P_n if k is odd. A similar result holds for Q_n.

Let $b_n = P_n/n!$ or $Q_n/n!$. The recurrence yields

$$b_n = b_{n-1} + \frac{b_{n-2}}{n(n-1)}.$$

Since $b_1, b_2 \geq 1$, we obtain by induction $1 \leq b_{n-1} < b_n$ for all $n \geq 3$. Thus

$$b_n < \left(1 + \frac{1}{n(n-1)}\right) b_{n-1} < b_2 \prod_{j=3}^{n} \left(1 + \frac{1}{j(j-1)}\right)$$

$$< b_2 \prod_{j=3}^{\infty} \left(1 + \frac{1}{j(j-1)}\right) < \infty,$$

since $\Sigma_j [j(j-1)]^{-1}$ converges. Therefore $b_\infty = \lim_{n\to\infty} b_n$ exists, and $b_n < b_\infty$ for all n.

Let $b_n = b_\infty(\Sigma_{i=0}^{k}(-1)^i(n-i)!/(i!n!) + R(n,k))$ for $k \leq n$, so that the desired inequality is $(-1)^k R(n,k) < 0$. We prove this by induction on k. Note that it is true for $k = 0$ because $b_n < b_\infty$. Assume that $(-1)^k R(j,k) < 0$ for all $j \geqslant k$. For $n \geqslant k+1$, we write

$$\frac{b_n - b_\infty}{b_\infty} = -\sum_{j=n}^{\infty} \frac{b_{j+1} - b_j}{b_\infty}$$

$$= -\sum_{j=n}^{\infty} \frac{b_{j-1}}{j(j+1)b_\infty} = -\sum_{j=n}^{\infty} \left\{ \sum_{i=0}^{k} (-1)^i \frac{(j-i-1)!}{i!(j+1)!} + \frac{R(j-1,k)}{j(j+1)} \right\}$$

$$= -\sum_{i=0}^{k} \frac{(-1)^i}{i!} \sum_{j=n}^{\infty} \frac{(j-i-1)!}{(j+1)!} - \sum_{j=n}^{\infty} \frac{R(j-1,k)}{j(j+1)}.$$

The interchange of summations is justified by absolute convergence. For fixed i, the sum $\Sigma_{j=n}^{\infty}(j-i-1)!/(j+1)!$ is telescoping, because

$$\frac{(j-i-1)!}{(j+1)!} = \frac{1}{i+1}\left(\frac{(j-i-1)!}{j!} - \frac{(j-1)!}{(j+1)!}\right).$$

Thus we obtain

$$\frac{b_n}{b_\infty} = 1 - \sum_{i=0}^{k} \frac{(-1)^i (n-i-1)!}{(i+1)!n!} - \sum_{j=n}^{\infty} \frac{R(j-1,k)}{j(j+1)}$$

$$= \sum_{i=0}^{k+1} \frac{(-1)^i (n-i)!}{i!n!} - \sum_{j=n}^{\infty} \frac{R(j-1,k)}{j(j+1)}.$$

If we replace b_n by the formula defining $R(n, k+1)$, namely

$$b_n = b_\infty \sum_{i=0}^{k+1} \frac{(-1)^i (n-i)!}{i!n!} + R(n, k+1),$$

we obtain

$$R(n, k+1) = -\sum_{j=n}^{\infty} \frac{R(j-1, k)}{j(j+1)}.$$

Hence the sign of $R(n, k+1)$ is opposite to the common sign of $R(n-1, k)$, $R(n, k), \ldots$, which completes the proof.

Solved also by P. Bracken (Canada), D. Callan, L. L. Foster, E. Grosswald, S. Heller, K.-W. Lau (Hong Kong), A. M. Mercer (Canada), W. A. Newcomb, C. S. Sutton, the Howard University Problems Group, and an anonymous contributor. R. K. Guy and Len Bos remark that the sequences $\{P_n\}$ and $\{Q_n\}$ are given as sequences 704 and 1151 in N. J. A. Sloane's *Handbook of Integer Sequences*, Academic Press, 1973.

Equilateral Triangles Inscribed in Closed Curves

E 3273 [1988, 555]. *Proposed by Orrin Frink, Pennsylvania State University.*

Give appropriate conditions under which a simple closed curve in the plane contains three points that form the vertices of an equilateral triangle. For the corresponding problem for the square, see Problem 4325 [1949, 39; 1950, 423] and Victor Klee, "Some unsolved problems in plane geometry", *Math. Magazine* 52(1979), 131–145.

Solution by Albert Nijenhuis, Seattle. No extra conditions are necessary. Let C be a simple closed curve with interior S and exterior S'. Given any points $P, Q \in C$, let $\phi(P, Q)$ be the third vertex of the equilateral triangle with base PQ that lies on the left of the ray PQ. If P, Q are chosen at maximum distance, then either $\phi(P, Q) \in C$, solving the problem, or $\phi(P, Q) \in S'$. Since ϕ is continuous, it suffices to show there exist P, Q such that $\phi(P, Q) \in S$. Choose $R \in S$, and obtain D by rotating C through 60 degrees about R. Since C and D contain the same area, neither can lie inside the other, so there is a point $Q \in C \cap D$. If $P \in C$ is its pre-image under the rotation, then $R = \phi(P, Q)$, as desired.

Editorial comment. Mark Meyerson points out that the problem was solved by A. N. Milgram in "Some metric topological invariants," *Reports of a Mathematical Colloquium* (Univ. of Notre Dame), second series 5-6(1939–1948), 25–39, where he also generalized the result to closed curves in any metric space. In "Equilateral triangles and continuous curves," *Fundamenta Mathematicae* 110(1980), 1–9, Meyerson shows that all but at most two points of a simple closed curve in the plane are vertices of such a triangle, and that for any fixed triangle Δ, every simple closed curve in the plane contains the vertices of a triangle similar to Δ. He also notes that the generalization of the square problem to three dimensions is false (MONTHLY Problem 6385 [1982, 279; 1983, 651]). Although Jerrard proved that a *smooth* simple closed curve in the plane has an inscribed square (*Trans. Amer. Math. Soc.*, 98(1961) 234–241), the existence of an inscribed square in a general simple closed curve in the plane is still open and is quite difficult. The solution printed for Problem 4235 [1950, 423] is not adequate.

Solved also by O. P. Lossers (The Netherlands), M. D. Meyerson, and Victor Pambuccian.

Sufficient Conditions for Ring Commutativity

E 3275 [1988, 555]. *Proposed by Thomas J. Laffey, University College, Dublin, and Desmond MacHale, University College, Cork, Ireland.*

(a) Suppose R is an associative ring, not necessarily with unity, in which every element is a square. If squaring is a ring homomorphism, prove that R is commutative.

(b) Suppose R is an associative ring and suppose there is a function f from R onto R which satisfies $f(x + y) = f(x) + f(y)$ and $f(xy) = (xy)^2$ for all $x, y \in R$. Must R be commutative?

Solution by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. (a) Note that $x^2 + y^2 = x^2 + xy + yx + y^2 = (x + y)^2$ if and only if $xy = -yx$. Thus squaring is an additive group homomorphism if and only if R is anti-commutative. If every element is a square, anticommutativity implies commutativity.

(b) R need not be commutative. An elementary counterexample is the set of all sequences of integers with the usual termwise addition, and with multiplication defined by

$$(x_1, x_2, x_3, \dots)(y_1, y_2, y_3, \dots) = (x_2 y_3, 0, 0, \dots).$$

This ring is not commutative, but the function f defined by $f(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots)$ has the specified properties.

Editorial comment. Many readers noted that the hypothesis of (a) can be weakened by deleting the requirement that f be a multiplicative homomorphism, and that the conclusion can be strengthened by adding that the ring has characteristic 2. Readers also noted that in (b) commutativity would be implied by a variety of additional hypotheses on R.

Complete solutions were submitted also by S. F. Barger, F. J. Flanigan, A. A. Jagers (The Netherlands), C. Lanski, and O. P. Lossers (The Netherlands). Part (a) was solved by 15 readers and the proposers.

Closed Sets in a Special Topological Space

E 3280 [1988, 655]. *Proposed by Peter J. Ferraro, Roselle Park, NJ.*

Let (X, T) be a topological space such that $\{p\}'$ is closed for every point $p \in X$. Prove that A' is closed for every subset A of X. Here A' is the set of points $x \in X$ such that every open set containing x contains a point of $A - \{x\}$.

Solution by José Heber Nieto, Universidad del Zulia, Maracaibo, Venezuela. We show that $A'' \subseteq A'$ for every $A \subset X$. Let z be a point in A'' and U be an open set containing z. Then $V = U - \{z\}'$ is also open and contains z, since $\{z\}'$ is closed and $z \notin \{z\}'$. Therefore, there is a point $x \in A'$ such that $x \in V - \{z\}$. Since $x \notin \{z\}'$, there is an open set W such that $x \in W$ and $z \notin W$. Then $W \cap V$ (and therefore U) contains a point of $A - \{z\}$.

Editorial comment. Many readers noted that the problem appears as Problem D in Chapter 1 of J. L. Kelley's *General Topology*, Van Nostrand, 1955. Kelley

attributes the problem to C. T. Yang. The result of the problem also appears as Theorem 14.5 in W. J. Thron's *Topological Structures*, Holt, Rinehart, and Winston, 1966. Peter Landweber notes that it is customary to call A' the *derived set* of A and its members the *accumulation points* of A, so that the statement becomes "the set of accumulation points of every subset is closed if and only if the set of accumulation points of each singleton set is closed." Stephan Carlson proved the stronger result that these properties are equivalent to a third property: Every singleton subset of X is the intersection of a closed set and an open set.

Solved also by the proposer and about 40 other readers.

Valence of Analytic Functions

E 3295 [1988, 953]. *Proposed by A. W. Goodman, University of South Florida, Tampa.*

A function f on the open unit disc $E = \{z: |z| < 1\}$ is said to be *univalent* if it is analytic on E and takes no value more than once. A function f on E is said to be *infinite-valent* if it is analytic on E and takes some value infinitely often.

(a) Is there a univalent function f on E such that f' is nonconstant and infinite-valent on E?

(b) Is there an infinite-valent function g on E such that g' is univalent on E?

Solution by Kee-Wai Lau, Hong Kong. We answer both questions in the affirmative. An elementary consideration of the range of the principal branch of $\log(1 + z)$ shows that $f(z) = (1 + z)^{1+i}$ is univalent, $f'(z) = g(z) = (1 + i)(1 + z)^i$ is infinite-valent, and $f''(z) = g'(z) = (-1 + i)(1 + z)^{i-1}$ is univalent. That f and g' are univalent is a special case of a lemma (A. W. Goodman, *Univalent Functions, Vol. II* (Polygonal Publishing Company, Inc., 1983), Chapter 15, Lemma 1), stating that for $\beta \neq 0$, the function $(1 + z)^\beta$ is univalent on E if and only if either $|\beta - 1| \leq 1$ or $|\beta + 1| \leq 1$. That f' and g are infinite-valent follows from the fact that

$$f'(-1 + e^{-2\pi n}) = g(-1 + e^{-2\pi n}) = (1 + i)e^{i(-2\pi n)} = 1 + i$$

for $n = 1, 2, 3, \ldots$.

Solved also by Wancang Ma and the proposer.

ADVANCED PROBLEMS

6622. *Proposed by Robert C. Dalang, École Polytechnique Fédérale, Lausanne, Switzerland.*

Does there exist a Lebesgue measurable function $f:[0,1] \to [0,1]$ which is positive on a set of positive measure and which satisfies

$$\sum_{n=1}^{\infty} \sum_{j=0}^{2^{n-1}-1} \left\{ \int_{2j/2^n}^{(2j+1)/2^n} f(x)\,dx \int_{(2j+1)/2^n}^{(2j+2)/2^n} f(x)\,dx \right\}^{1/2} < +\infty?$$

6623. *Proposed by Ernesto Bruno Cossi, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, and the editors.*

Let $R(n)$ denote the number of ways of expressing the positive integer n as a sum of two squarefree positive integers relatively prime to n. Is it true that $R(n) > c\,\phi(n)$ for some positive constant c, where ϕ denotes the Euler function?

6624. *Proposed by K. A. Ross, University of Oregon, Eugene.*

(i) Show that there is a subset A of the plane R^2 that intersects every straight line in exactly two points.

(ii) Can the set A be selected to be Lebesgue measurable?

(iii)* Can the set A be selected to be a Borel set?

SOLUTIONS OF ADVANCED PROBLEMS

Vanishing "Triangular" Functions

6526 [1986, 659]. *Proposed by Grzegorz Rzadkowski, Agricultural University in Warsaw, Poland.*

Let f be a continuous real-valued function defined on the interval $[0, L]$. If there is an equilateral triangle ABC of sidelength L such that

$$f(AP) + f(BP) + f(CP) = 0$$

for all points P inside the triangle, prove that f is identically zero.

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Without loss of generality we take $L = 1$. It is easy to give a noncontinuous function on $[0,1]$ that has the desired property for all points P in the closed triangle: define $f(0) = 1$, $f(1) = -\frac{1}{2}$, and $f(x) = 0$ for $0 < x < 1$. We shall prove the following generalization of the assertion of the problem.

THEOREM. *Let f be any function defined on the open interval $(0,1)$ such that*

$$f(A_1P) + f(A_2P) + f(A_3P) = 0 \tag{1}$$

for all points P inside the equilateral triangle $A_1A_2A_3$ of side 1. Then f is identically zero (and hence continuous).

Proof. Let M be the centre of mass of the triangle. Then all vectors $\overrightarrow{MA}_k$ have length $(1/3)\sqrt{3}$ and subtend angles $2\pi/3$. So, by the cosine rule, we have

$$A_kP^2 = \tfrac{1}{3} + MP^2 - \tfrac{2}{3}\sqrt{3}\,MP\cos\left(\varphi + \frac{2k\pi}{3}\right) \qquad (k = 1,2,3)$$

for some φ. Elimination of MP and φ yields

$$\sum_k A_kP^4 - \sum_{k<l} A_kP^2 \cdot A_lP^2 - \sum_k A_kP^2 + 1 = 0. \tag{2}$$

It is useful to indicate an interior point P of the triangle by its triple of squares

$$P\colon (\pi_1, \pi_2, \pi_3) := \left(A_1P^2, A_2P^2, A_3P^2\right). \tag{3}$$

For example, M is indicated by M: $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. In general, we have by (2) for any P the relation

$$\pi_1^2 + \pi_2^2 + \pi_3^2 - \pi_1\pi_2 - \pi_1\pi_3 - \pi_2\pi_3 - \pi_1 - \pi_2 - \pi_3 + 1 = 0. \tag{4}$$

Consider any ordered 4-tuple of points, $[P, Q, R, S]$ such that $A_1Q = A_1R$, $A_2P = A_2R$, $A_2Q = A_2S$, $A_3P = A_3Q$, and $A_3R = A_3S$. (See Fig. 1). Such a 4-tuple will

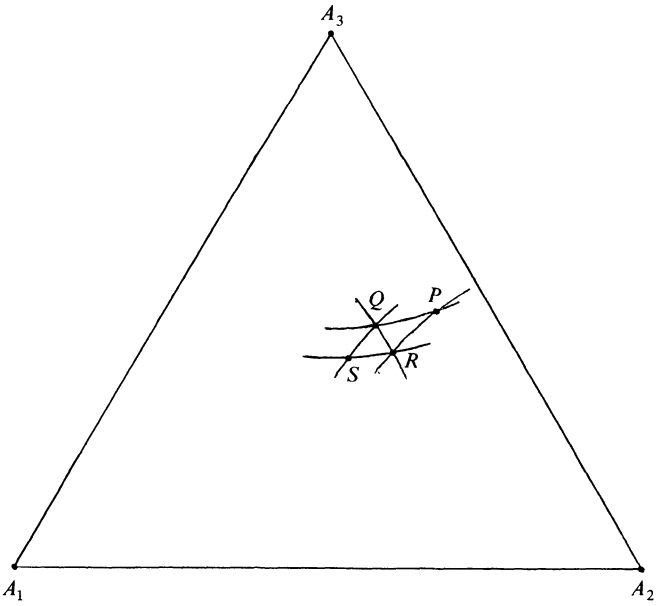


FIG. 1

be called "pentacyclic." Then, using (1), we find by elimination that

$$f(A_1P) - 2f(A_1Q) + f(A_1S) = 0. \tag{5}$$

A simple, important consequence of (5) is the fact that $f(x)$ is constant (and hence 0) when it is constant on some interval.

Such an interval will be given in the sequel.

We shall construct a continuously variable pentacyclic 4-tuple $[P, Q, R, S](t)$ $(0 \leqslant t \leqslant 1)$ such that $P(t)$ and $Q(t)$ (and hence $R(t)$) move along circles with centre A_1, but such that the distance $A_1S(t)$ is non-constant (and therefore runs through an interval of positive length, whereas $f(A_1S)$ remains constant).

Let $[P, Q, R, S]$ and $[T, M, U, V]$ be two special pentacyclic 4-tuples: $A_2P = A_3P$, $A_1Q = A_1M$ $(= (1/3)\sqrt{3})$, and $A_1P = A_1T$ (see Figure 2).

Assume that $A_1S = A_1V$.

By symmetry, the triple square coordinates of P, Q, R and S can be written in the form

$$\begin{cases} Q\colon \left(\frac{1}{3}, \frac{1}{3} + \alpha, \frac{1}{3} - \beta\right),\ R\colon \left(\frac{1}{3}, \frac{1}{3} - \beta, \frac{1}{3} + \alpha\right) \\ P\colon \left(\frac{1}{3} + \gamma, \frac{1}{3} - \beta, \frac{1}{3} - \beta\right),\ S\colon \left(\frac{1}{3} - \delta, \frac{1}{3} + \alpha, \frac{1}{3} + \alpha\right). \end{cases} \tag{6}$$

Also by symmetry we have

$$A_3V = A_3U = A_1T = A_1P.$$

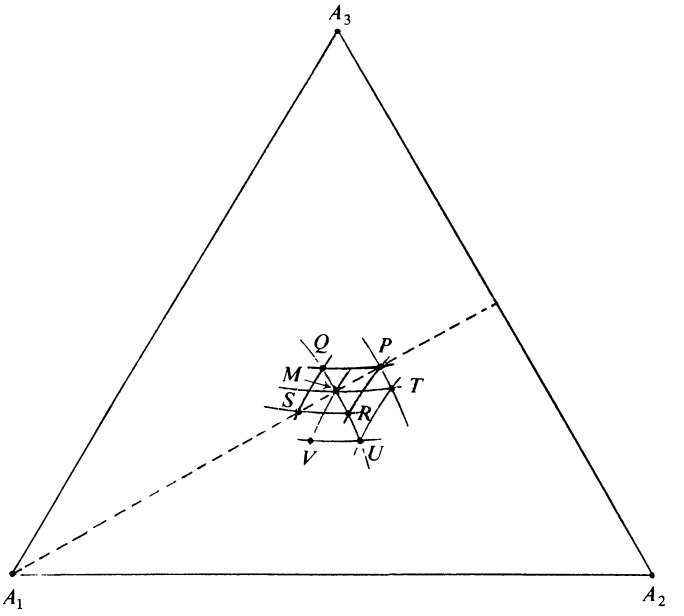


FIG. 2

This implies the following triple square coordinates for T, U, and V:

$$T\colon \left(\tfrac{1}{3}+\gamma, *, \tfrac{1}{3}\right),\ U\colon \left(\tfrac{1}{3}, *, \tfrac{1}{3}+\gamma\right),\ V\colon \left(\tfrac{1}{3}-\delta, \tfrac{1}{3}, \tfrac{1}{3}+\gamma\right). \tag{7}$$

Here the missing coordinates $*$ both have to be equal to $1/3 - \delta$, since two coordinates of a point uniquely determine the third coordinate. All points given by (6) and (7) must satisfy (4), so

$$\begin{cases} \alpha^2 + \alpha\beta + \beta^2 - \alpha + \beta = 0 & \text{(from } Q \text{ and } R\text{)} \\ \beta^2 + 2\beta\gamma + \gamma^2 + 2\beta - \gamma = 0 & \text{(from } P\text{)} \\ \gamma^2 + \gamma\delta + \delta^2 - \gamma + \delta = 0 & \text{(from } T, U \text{ and } V\text{)} \\ \delta^2 + 2\delta\alpha + \alpha^2 + \delta - 2\alpha = 0 & \text{(from } S\text{)}. \end{cases} \tag{8}$$

We are interested in solutions $(\alpha, \beta, \gamma, \delta) \neq (0, 0, 0, 0)$. Elimination of β and δ from (8) forces two resultants $R_1(\alpha, \gamma)$ and $R_2(\alpha, \gamma)$ to be zero:

$$\begin{aligned} R_1(\alpha,\gamma) &:= \begin{vmatrix} 1 & \alpha+1 & \alpha^2-\alpha & 0 \\ 0 & 1 & \alpha+1 & \alpha^2-\alpha \\ 1 & 2\gamma+2 & \gamma^2-\gamma & 0 \\ 0 & 1 & 2\gamma+2 & \gamma^2-\gamma \end{vmatrix}; \\ R_2(\alpha,\gamma) &:= \begin{vmatrix} 1 & \gamma+1 & \gamma^2-\gamma & 0 \\ 0 & 1 & \gamma+1 & \gamma^2-\gamma \\ 1 & 2\alpha+1 & \alpha^2-2\alpha & 0 \\ 0 & 1 & 2\alpha+1 & \alpha^2-2\alpha \end{vmatrix}. \end{aligned} \tag{9}$$

It is immediately verified that both $R_i(\alpha, \gamma)$ are quartic forms in α and γ together, and that

$$R_1(1, \gamma) = \gamma^2(\gamma - 1)(\gamma - 5),$$

whereas

$$R_2(1, \gamma) = \gamma^4 - 5\gamma^3 + 11\gamma^2 - 7\gamma + 3.$$

Thus the four roots of $R_1(1, \gamma) = 0$ are all distinct from the four roots of $R_2(1, \gamma) = 0$, i.e., the quartics $R_1(\alpha, \gamma) = 0$ and $R_2(\alpha, \gamma) = 0$ do not have a component in common.

In other words $(\alpha, \beta, \gamma, \delta) = (0, 0, 0, 0)$ is an isolated solution of (8), so our assumption that $A_1S = A_1V$ is false, provided that P is sufficiently close to M.

Therefore, for P sufficiently close to M, the pentacyclic 4-tuples $[P, Q, R, S]$ and $[T, M, U, V]$ can be used as the endpoints of a continuously variable 4-tuple $[P, Q, R, S](t)$, where $0 \leqslant t \leqslant 1$. The result follows.

Also solved by the San Bernardino Problem Solving Group, Aage Bondesen (Denmark), and David C. Kay.

The Harmonic Conjugate of an Algebraic Function

6567 [1988, 150]. *Proposed by Gerald A. Edgar, Ohio State University, and Lee A. Rubel, University of Illinois at Urbana-Champaign.*

Let us say that a function $u(x, y)$ is algebraic if there exists a nontrivial polynomial P in three variables such that $P(x, y, u(x, y)) = 0$. Suppose $u(x, y)$ is a harmonic and algebraic function on a region G in $\mathbb{R}^2$. Must its harmonic conjugate v also be algebraic?

Solution by the proposers. Yes, v must be algebraic. First we show that $f(x + iy) = u(x, y) + iv(x, y)$ is algebraic as a function of $z = x + iy$. Write $\bar{z} = x - iy$ and

$$\tilde{u}(z, \bar{z}) = u(x, y), \qquad \tilde{v}(z, \bar{z}) = v(x, y).$$

If $f(z) = \sum_{n=1}^{\infty} a_n z^n$ (suppose for convenience that $(0, 0) \in G$ and that $f(0) = 0$) then

$$\tilde{f}(z) = \sum \bar{a}_n \bar{z}^n$$

and

$$\tilde{u}(z, \bar{z}) = \frac{1}{2}\left[\sum a_n z^n + \sum \bar{a}_n \bar{z}^n\right].$$

Now for some nontrivial polynomial Q, we have

$$Q\left(z, \bar{z}, \sum a_n z^n + \sum \bar{a}_n \bar{z}^n\right) = 0. \qquad (*)$$

Suppose for simplicity that $Q(z, \bar{z}, w)$ has no factor $\bar{z}^k$, $k > 0$. In $(*)$, set $\bar{z} = 0$ to get $Q(z, 0, f(z)) = 0$. Hence $f(z)$ is algebraic, as claimed.

We need some explanation of the above directions "set $\bar{z} = 0$." It is well known that $u(x, y)$ is represented by a convergent series of powers of x and y near $(0, 0)$. Thus we may consider x and y as *complex* variables, and then consider $u(x, y)$ on the set $\{x = iy\}$.

We conclude, from the just-established fact that $f(z)$ is an algebraic function of z, that it is also an algebraic function of $(z, \bar{z})$. From this it follows easily that $\tilde{f}(z)$

is an algebraic function of $(z, \bar{z})$. But, since the difference of any two algebraic functions is algebraic, we see that

$$2i\tilde{v}(z, \bar{z}) = f(z) - f(\bar{z})$$

is an algebraic function of $(z, \bar{z})$, and consequently that $\tilde{v}(z, \bar{z})$ is an algebraic function of $(z, \bar{z})$. Hence $v(x, y)$ is an algebraic function of (x, y).

No other solution was received.

Infinite Cardinals

6569 [1988, 264]. *Proposed by Wilson Castrellon and Samuel Camargo, Universidad de Los Andes, Bogotá, Colombia.*

Show that for every infinite cardinal λ there exists a cardinal $\kappa \geqslant \lambda$, a set A of cardinality κ, and a family F of subsets of A such that $\operatorname{card}(F) = 2^{\kappa}$ and $X \cap Y$ is finite for all $X, Y \in F$, $X \neq Y$.

Solution by Fred Galvin, University of Kansas, Lawrence. We first claim there is a cardinal $\kappa > \lambda$ such that

$$\kappa^{\aleph_0} = 2^{\kappa}.$$

Choose

$$\kappa = \lambda + 2^{\lambda} + 2^{2^{\lambda}} + 2^{2^{2^{\lambda}}} + \cdots .$$

By routine cardinal arithmetic,

$$\kappa^{\aleph_0} \leqslant (2^{\kappa})^{\aleph_0} = 2^{\kappa \cdot \aleph_0} = 2^{\kappa}$$

and

$$2^{\kappa} = 2^{\lambda} \cdot 2^{2^{\lambda}} \cdot 2^{2^{2^{\lambda}}} \cdots \leqslant \kappa \cdot \kappa \cdot \kappa \cdots = \kappa^{\aleph_0}.$$

The claim follows. Now let S be a set of cardinality κ, let A be the set of all finite sequences of elements of S, and B the set of all infinite such sequences. Each β in B corresponds in a natural way to a subset of A, namely the set $A(\beta)$ of all initial segments of β. In other words, if

$$\beta = (x_1, x_2, x_3, \ldots) \in B$$

then

$$A(\beta) = \{(x_1), (x_1, x_2), (x_1, x_2, x_3), \ldots\} \subseteq A.$$

If

$$F = \{A(\beta) : \beta \in B\},$$

then

$$\operatorname{card}(F) = \kappa^{\aleph_0}$$

and κ has been chosen so that this is 2^{κ}. Finally, $\beta_1 \neq \beta_2$ implies

$$A(\beta_1) \cap A(\beta_2)$$

is finite.

All solutions received were essentially the same; that of Dennis Burke described a construction of the above type in the language of topology.

Also solved by Andreas Blass, Charles Lanski, James C. Owings, Jr. & Philip W. Steitz, István Szalkai (Hungary), and F. P. Weber (Canada).

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

*Representations of *-Algebras, Locally Compact Groups, and Banach *-Algebraic Bundles, I, II.* By J. M. G. Fell and R. S. Doran, Academic Press, 1988, xviii + 1486 pp.

MARC A. RIEFFEL
Department of Mathematics, University of California, Berkeley, CA 94720

Groups provide the language for describing symmetries. In this role they act as groups of automorphisms of mathematical objects. (One thinks of Galois groups in algebra, of Lie groups in geometry, and of the many symmetry groups of physics). Accordingly, it is important to understand the structure of actions of groups. Foremost actions are those on vector spaces (that is, linear actions, or "representations," as they are more commonly called), because any action can be linearized, by considering the corresponding action by translation on the vector space of scalar-valued functions on the object. There is by now a vast literature concerned with representations.

Frobenius, Schur, and others first studied finite groups acting on finite-dimensional vector spaces around the beginning of this century. This theory is by now highly developed. Soon Weyl and others saw that the theory of continuous representations of compact groups, such as the group of rotations of Euclidean space, is quite similar to that for finite groups; and so this theory too is now well understood. But for noncompact groups (even discrete ones) a sharp dichotomy exists. On the one hand, there are the representations by unitary operators on Hilbert spaces. These have the crucial property that if a subspace of the Hilbert space is invariant (that is, carried into itself by the actions of all the elements of the group), then so is its orthogonal complement, and the original representation is then the direct sum of these two subrepresentations. Thus one can hope that unitary representations can be decomposed in some relatively transparent way into the "simple" pieces consisting of the irreducible representations (those with no invariant subspaces), so that the theory is "semisimple." The main task is then to determine the irreducible representations.

On the other hand, if one considers nonunitary representations, then even when one can determine the irreducible representations, one is still left with the often bewildering problem of understanding all the ways in which irreducible representations can be combined to obtain more general representations. For the group Z of integers and its representations on finite-dimensional complex vector spaces, this leads to the Jordan canonical form for matrices; while for the groups Z^n for $n \geqslant 3$ I believe that the situation is still poorly understood. For this reason, while at present the theory of unitary representations is very well developed (even though for many groups it is essential that one work with infinite-dimensional Hilbert spaces), the theory of nonunitary representations (even just on finite-dimensional vector spaces) consists of relatively fragmentary pieces, with few unifying principles.

Even for unitary representations, the theory is well developed only for those groups whose topology is locally compact. The basic reason for this is that for groups which are not locally compact one is often hard pressed to exhibit any nontrivial, continuous unitary representations at all. But for locally compact groups Haar, in 1933, showed that they possess an analogue of Lebesgue measure, namely, a positive Borel measure which is invariant under left translation; then one sees that any locally compact group actually has a faithful unitary representation, consisting of translation on the Hilbert space of functions on the group which are square-integrable for this Haar measure. A few years later, Weil showed that locally compact groups are the only ones to possess such a Haar measure. Nevertheless, more recently the needs of various subjects have led to much discussion of the unitary representation theory of certain classes of groups which are not locally compact (loop groups, gauge groups, diffeomorphism groups—see [5] and the references it contains). But this theory is still in its infancy, with few general results and unifying principles.

The literature concerning unitary representations of locally compact groups splits to some extent into two parts. The larger part consists of treatments of particular classes of groups (say, nilpotent Lie groups), with the aim of obtaining explicit classification and construction of their irreducible representations, or of describing explicitly how various naturally occurring representations decompose into irreducible representations. This subject can be extremely rich in detail (for example, in the case of semisimple Lie groups).

The second, and smaller, part of the literature consists of discussions of general methods and principles. Perhaps the first striking result in this category was the demonstration by Gelfand and Raikov in 1943 that every locally compact group has a separating family of irreducible unitary representations (with a necessarily nonconstructive proof, since so little is assumed about the structure of the group). This general theory is now highly developed. It receives its first thorough modern exposition in the two-volume work under review (though [1] comes somewhat close, and contains a broad collection of examples and topics).

The two volumes differ somewhat in character, with the first volume, entitled *Basic Representation Theory of Groups and Algebras*, being devoted to the more foundational material. In the theory of representations of finite groups, one of the first steps is to introduce the group algebra of the group, consisting of the scalar-valued functions on the group, with convolution as product. This is useful because the representations of the group correspond exactly to the representations of the algebra, while the two operations in an algebra allow much more flexibility of manipulation than one has with the group alone. When one studies representations of locally compact groups, it is again useful to introduce an appropriate group algebra having the universal property that its representations correspond exactly to the continuous unitary representations of the group. This consists of the continuous functions of compact support with convolution as product (making another essential use of Haar measure), with an involution corresponding to the map $x \mapsto x^{-1}$ at the group level, and then completed with respect to a norm which comes from the unitary representations. The result is a C^*-algebra, that is, a Banach algebra over the complex numbers, with involution, whose norm satisfies the innocent-looking identity $\|a^*a\| = \|a\|^2$. (This identity has remarkable consequences; notably, according to a famous theorem of Gelfand and Naimark from 1943, it exactly

characterizes the Banach *-algebras which have faithful representations as norm-closed self-adjoint algebras of operators on a Hilbert space.) One is thus led to the study of C^*-algebras and their representations. Such study is more generally useful, because C^*-algebras arise in a number of related contexts, ranging from quantum statistical mechanics, to foliations of manifolds, to dynamical systems. The first volume consists of a thorough and leisurely exposition both of the basic theory of locally compact groups, and of C^*-algebras and their representations, with emphasis on those aspects most germaine to the unitary representation theory of groups.

The second volume is more specialized in character. The most important general construction in the representation theory of groups is that of induced representations. This construction takes a representation of a subgroup of a group, and constructs from it a representation of the group itself. For finite groups, if one views matters in terms of group algebras, this is nothing more than a familiar change-of-rings construction, by which one takes a module over a subalgebra and constructs from it a module over the algebra. Induced representations for finite groups were introduced by Frobenius at the turn of the century (see [3] for a nice description of the history); while for locally compact groups the general theory of induced representations was developed by Mackey in a series of influential papers in the 1950s (though Wigner's 1939 paper on the representation theory of the Poincaré group was a vital precursor).

The theory of induced representations gives its most definitive results when one is analyzing what happens in the presence of a normal subgroup, N, of the group in question, say, H. Let $G = H/N$. Then the action of H by conjugation on N gives an action of G on the isomorphism classes of irreducible representations of N. If this action is sufficiently well-behaved, there is an algorithm, using induced representations, for constructing the irreducible representations of H in terms of those of N and of the representations of the stability subgroups in G for the isomorphism classes of irreducible representations of N. This algorithm is often fondly referred to as "the Mackey machine."

The second volume is primarily devoted to the theory of induced representations, culminating in a thorough treatment of the Mackey machine. Actually, the senior author, in a series of earlier research papers, had shown that it is very natural to develop the Mackey machine in a more general context, in which the quotient group G above is retained, but H and N are replaced by a more general structure called a Banach *-algebraic bundle over G. It is in this more general context that the authors develop the Mackey machine, hence the second volume's title, *Banach *-Algebraic Bundles, Induced Representations, and the Generalized Mackey Analysis*.

Actually, if one looks at the extensive literature concerned with classifying the irreducible representations of specific classes of groups, one finds that much of it does not mention C^*-algebras at all (though various kinds of induced representations are ubiquitous). The reason is that there is yet another dichotomy in the subject. Many groups have a quite tractable unitary representation theory. For historical reasons they are said to be "of type I." (Nilpotent and semisimple Lie groups are of type I.) The theory of C^*-algebras is not essential for studying the unitary representation theory of type I groups (though it still can provide a very convenient framework). But groups not of type I have a badly behaved representation theory—there is no middle way. (Many solvable Lie groups, and most infinite discrete groups are not of type I.) There is ample evidence that the irreducible

representations of nontype-I groups are essentially unclassifiable. Accordingly, there is little literature concerning their representation theory and, in recent years, as the needs of other subjects have made it increasingly important to deal with nontype-I groups, attention has necessarily shifted away from their representation theory and towards the group C^*-algebra itself. By now, this has permitted a number of important advances, ranging from Pukansky's treatment of nontype-I Lie groups over a decade ago [**6**], to the very recent successes with the Novikov conjecture [**4**, **2**], where the effects of the fundamental group of a manifold are controlled by using its group C^*-algebra. The two volumes do not stress the nontype-I aspects (although examples are given), but the modern treatment which they provide should prove useful in the further development of the nontype-I theory.

Finally, let me comment that these volumes have been prepared with great care. The exposition is clear and thorough. At the end of each chapter there is an extensive collection of exercises and a brief historical summary, and each volume concludes with an extensive bibliography, name index, subject index, and index of notation.

REFERENCES

1. A. O. Barut and R. Raczka, Theory of Group Representations and Applications, Polish Scientific Publishers, 1977.
2. A. Connes and H. Moscovici, Conjecture de Novikov et Groupes Hyperboliques, preprint.
3. T. Hawkins, New light on Frobenius' creation of the theory of group characters, *Arch. History Exact Sci.*, 12 (1974) 217–243.
4. G. G. Kasparov, Equivariant KK-theory and the Novikov conjecture, *Invent. Math.*, 91 (1988) 147–201.
5. A. Pressley and G. Segal, Loop Groups, Oxford, 1986.
6. L. Pukansky, Characters of connected Lie groups, *Acta Math.*, 133 (1974) 81–137.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S, P.** *Speaking Mathematically: Communication in Mathematics Classrooms.* David Pimm. Routledge, Chapman & Hall, 1987, xvii + 217 pp, $14.95 (P). [ISBN: 0-415-03708-5] Very interesting study of how students perceive mathematics, especially its language. The primary thesis is that students cannot think effectively (in mathematics) until they have developed what the author calls a "mathematical register," that is, a reservoir of terms with their meanings developed in some sort of context. The author claims, and documents, that many if not most students learn about the form but not the meaning of mathematics. This book should be of interest to anyone who is wondering what students are learning in their classroom. MW

General, P, L*. *Shiing-shen Chern: Selected Papers, Volume II-IV.* Shiing-shen Chern. Springer-Verlag, 1989. *Volume II,* xxx + 444 pp, $45 [ISBN: 0-387-96816-4]; *Volume III,* xiv + 504 pp, $49 [ISBN: 0-387-96817-2]; *Volume IV,* xiv + 462 pp, $45. [ISBN: 0-387-96820-2] A complement to *Volume I* (TR, March 1979) which contained a selection of shorter and less-accessible papers, these three volumes contain major papers arranged chronologically: 1935-1965, 1965-1981, 1981-1987. Includes a complete Chern bibliography and (in *Volume II*) a recently revised scientific autobiography of Chern. LAS

General, P*. *Symbolism: Sculptures and Tapestries by John Robinson.* Ronnie Brown, John Robinson. Mathematics and Knots (School of Mathematics, U. of Wales, Bangor, Gwynedd LL57 1UT, Wales, UK), 1989, 36 pp, £6 + £1 surface, £250 air (P). [ISBN: 0-9514947-08] An exhibit catalogue prepared by mathematician Ronnie Brown to accompany a display of mathematically-inspired sculptures by John Robinson that have been assembled for the Pop Maths Road Show that opened in September 1989 in conjunction with the ICME workshop at the University of Leeds on popularization of mathematics. Trefoil knots, Möbius strips, tori, and other mathematical forms come to life in four-foot bronzes. LAS

General, S*, L*.** *The Penguin Dictionary of Mathematics.* Eds: John Daintith, R.D. Nelson. Penguin Books, 1989, 350 pp, $7.95 (P). [ISBN: 0-14-051119-9] A student's delight—virtually every term a college mathematics student would be expected to know (and lots more too) in a compact 8 oz. paperback. 2800 definitions, including 200 one-paragraph biographies. Extensive cross references signalled by asterisks by keywords helps reduce repetition. No computer terms, but does cover basic statistics, algebra, analysis, logic, topology, number theory, and other subjects from school and college mathematics. LAS

General, S. *NYSML-ARML Contests, 1983-1988.* Gilbert Kessler, Lawrence Zimmerman. NCTM, 1989, x + 148 pp, $18 (P). [ISBN: 0-87353-278-3] Problems and solutions from annual high school mathematics competitions, perhaps the most prestigious of the high school contests. A reader cannot help but be caught up in the excitement and intellectual energy represented in these high-quality problems. Readers at all levels can use these problems to pique their creative instincts. A valuable resource which has much to offer toward developing excellence in mathematics at the high school level. A useful Index covers all fifteen years of the contests. LCL

General, L*. *Izrail M. Gelfand: Collected Papers, Volume III.* Ed: S.G. Gindikin, *et al.* Springer-Verlag, 1989, x + 1075 pp, $149.50. [ISBN: 0-387-19399-5] Final installment in the Gelfand *Selecta (Volume I,* TR, December 1988; *Volume II,* TR, March 1989). This volume features papers on integral geometry, cohomology, and hypergeometric functions. It also includes samples of Gelfand's many research papers in cell biology (not mathematical models, but biological research), and some work in information theory and cybernetics. LAS

General, S*, L.** *Mathematical Carnival.* Martin Gardner. MAA, 1989, 297 pp, $15. [ISBN: 0-88385-448-1] Reprint of a 1975 Alfred A. Knopf anthology of *Scientific American* columns with a new introduction by John Conway, a new preface by Martin Gardner, and recent updates. This sixth of Gardner's collected columns is the first of several to be reprinted

by the M.A.A. Topics include sprouts, superellipses, trisection, Escher, and many puzzles. A superb classroom supplement to awaken bored students. LAS

Elementary, T. *Mathematics: Its Power and Utility, Third Edition.* Karl J. Smith. Brooks/Cole, 1989, xiii + 578 pp. [ISBN: 0-534-11538-1] Gives student "who has previously not been successful with mathematics a fresh and innovative approach to arithmetic, beginning algebra, and geometry." The Preface gives no indication of changes made in this edition. (*First Edition,* TR, June-July 1983; *Second Edition,* TR, November 1986.) JNC

Precalculus, T*(13: 1). *College Algebra and Trigonometry: A Graphing Approach.* Franklin Demana, Bert K. Waits. Addison-Wesley, 1990, xvi + 907 pp, $39.75 [ISBN: 0-201-52810-X]; *Graphing Calculator and Computer Graphing Laboratory Manual: Precalculus Series.* Alan Osborne, Gregory D. Foley. Addison-Wesley, 1990, 116 pp, $6.50 (P). [ISBN: 0-201-50861-3] New interactive instructional approaches permit a focus on problem solving themes such as business, economics, and science. Traditional algebraic topics arise from the context of mathematical models. Uses computer and calculator-based technology to involve students in conjecturing, hypothesis testing, and data analysis. The text and lab manual provide the resources for implementing needed changes in mathematics education. KT

Education, S, P**, L**.** *Historical Topics for the Mathematics Classroom.* John K. Baumgart. NCTM, 1989, xv + 542 pp, $28 (P). [ISBN: 0-87353-281-3] A collection of classroom capsules on the history and development of mathematics. Includes chapters on computation, algebra, geometry, trigonometry, calculus and "modern" mathematics. Originally published in 1969 as the *Thirty-first Yearbook* of the NCTM. Changes from the original edition included updated bibliographic references and a new chapter on contemporary mathematics. All teachers of mathematics at all levels should have this book (students too!). SB

History, T*(15-16), S*, P, L*. *Mathematics and Its History.* John Stillwell. Undergrad. Texts in Math. Springer-Verlag, 1989, x + 371 pp, $49.80. [ISBN: 0-387-96981-0] Written from notes for a senior-level course, this book "aims to give a unified view of undergraduate mathematics by approaching the subject through its history." Mathematical exercises throughout, ranging from computational to theoretical. Topics: Greek geometry and number theory, infinity, polynomial equations, analytic and projective geometry, calculus, infinite series, elliptic functions, mechanics, complex numbers in algebra, complex functions, differential and non-Euclidean geometry, group theory, topology, and sets, logic and computation. Excellent bibliography. Assumes basic calculus, algebra, and geometry as well as some exposure to group theory, topology, and differential equations. SB

History, S, L*. *Makers of Mathematics.* Stuart Hollingdale. Penguin Books, 1989, xiv + 433 pp, $10.95 (P). [ISBN: 0-14-022732-6] A selective, idiosyncratic history of mathematics told through a series of sketches of the lives and accomplishments of famous *men* (no women) of mathematics: Euclid, Apollonius, Archimedes, Descartes, Fermat, Pascal, Newton, Leibniz, Euler, D'Alembert, Gauss, Hamilton, Boole, Dedekind, Cantor, Einstein. Based in part on a series of biographical essays published by the author in the British IMA *Bulletin,* the volume has been carefully edited for a general reader with no more than a good school background in mathematics. A good mix of people and ideas that will appeal to many students. LAS

History, P, L.** *A Century of Mathematics in America, Part III.* Ed: Peter Duren. History of Math., V. 3. AMS, 1989, ix + 675 pp, $75. [ISBN: 0-8218-0136-8] Third and final volume containing historical papers written in honor of the AMS centennial. This volume contains 35 papers including departmental histories (Hopkins, Clark, IAS, Columbia, MIT, Michigan, Texas); personal reminiscences; topical histories; historical perspectives on probability, statistics, and actuarial science; and the history of the history of mathematics in America. Includes an informative analysis of American calculus textbooks of the nineteenth century. (*Part I* and *II,* TR, October 1989.) LAS

History, T(13-16: 1), L. *An Introduction to the History of Mathematics with Cultural Connections, Sixth Edition.* Howard Eves. Saunders College, 1990, xviii + 775 pp, $38. [ISBN: 0-03-029558-0] Changes from the previous edition (TR, October 1983) include updatings, some new sections, expansion of some old sections, more attention to women in mathematics, and an improved bibliography. Also added are Cultural Connections written by Jamie H. Eves, which discuss the cultural setting of various eras in the history of mathematics. (*Third Edition,* TR, October 1969; *Fourth Edition,* TR, August-September 1976.) RH

History, L*. *Introduction to Analysis of the Infinite, Book II.* Leonard Euler. Transl: John D. Blanton. Springer-Verlag, 1990, xii + 504 pp, $59. [ISBN: 0-387-97132-7] The theory of curves (or "higher-geometry"), including classification of algebraic curves, curvature, intersections, tangencies, and transcendental curves. Includes an Appendix on surfaces. The second half of Euler's *Introductio in analysin infinitorum,* here translated into English for the first time (for review of *Book I,* see TR, February 1989). After 240 years, retains an amazing clarity of language and notation. LAS

Logic, P. *Computability in Analysis and Physics.* Marian B. Pour-El, J. Ian Richards. Perspect. in Math. Logic. Springer-Verlag, 1989, xi + 206 pp, $65. [ISBN: 0-387-50035-9] Quantum theorists take note. There's a tendency to think that once you know the operator, you can automatically compute the eigenvalues and eigenvectors—but it ain't necessarily so. Modern mathematical logic has its say

in the matter: "With certain mild side-conditions, bounded operators preserve computability and unbounded operators do not." Moreover, even though the eigenvalues of a self-adjoint operator are computable, the *sequence* of all eigenvalues need not be computable. BC

Graph Theory, T(16-18), S, P, L. *Hypergraphs: Combinatorics of Finite Sets.* Claude Berge. Math. Lib., V. 45. North-Holland (US Distr: Elsevier Science), 1989, x + 255 pp, $79. [ISBN: 0-444-87489-5] The theory of hypergraphs is a generalization of graph theory wherein edges (e.g., pairs of vertices) are replaced with arbitrary sets of vertices. This generalization often leads to simplification, economy, and unification in understanding theorems on graphs. This book presents the most significant work, including many open problems. Topics include general concepts, transversal sets and matchings, fractional transversals, colorings, and generalizations of bipartite graphs. LCL

Combinatorics, P. *Surveys in Combinatorics, 1989.* Ed: Johannes Siemons. London Math. Soc. Lect. Note Ser., V. 141. Cambridge University Pr, 1989, viii + 217 pp, $27.95 (P). [ISBN: 0-521-37823-0] Papers by the principle speakers at the Twelfth British Combinatorial Conference, Norwich, April 1989. LC

Discrete Mathematics, T(13-14: 1), L. *Learning Discrete Mathematics with ISETL.* Nancy Baxter, Ed Dubinsky, Gary Levin. Springer-Verlag, 1989, xvii + 416 pp, $45. [ISBN: 0-387-96898-9] Students learn discrete mathematics by programming in ISETL (Interactive **SET** Language) which has syntax resembling mathematical language. Intent is more student participation (reading text on own, programming, solving problems), and less lecturing (spoon-feeding) by teacher. LC

Discrete Mathematics, L*, P. *Discrete Mathematics in the First Two Years.* Ed: Anthony Ralston. MAA Notes No. 15. MAA, 1989, 101 pp, $10 (P). [ISBN: 0-88385-064-8] Reports from six campuses—Colby, Delaware, Denver, Florida State, Montclair State, St. Olaf—that were funded by the Sloan Foundation in 1984-86 to experiment with various ways of incorporating discrete mathematics into the standard first two-year curriculum. Opens with an essay by Tony Ralston—the godfather of the discrete mathematics movement—and concludes with a list of discrete mathematics texts and other recommendations from the MAA Committee on Discrete Mathematics. An excellent resource for any department seeking a proper role for discrete mathematics. LAS

Number Theory, P. *Lecture Notes in Mathematics-1383: Number Theory.* D.V. Chudnovsky, *et al.* Springer-Verlag, 1989, 256 pp, $24.30 (P). [ISBN: 0-387-51549-6] Wide-ranging papers from the weekly number theory seminar held at the City University of New York. This, the fourth such volume, comes from talks given in 1985-1988. GG

Number Theory, T*(13-15: 2), S*, L*. *Introduction to Number Theory with Computing.* R.B.J.T. Allenby, E.J. Redfern. Edward Arnold, 1989, x + 310 pp, $22.95 (P). [ISBN: 0-7131-3661-8] Well-written introduction to elementary number theory which includes many computing applications. Contains numerous Basic programs and assignments to write your own. Lots of exercises. Does not, however, replace rigor by calculations (e.g., proves the fundamental theorem of arithmetic for $Z[i]$). Each chapter includes one or two short historical sketches (Fibonacci, Hilbert, Dirachlet, Mersenne, Fermat, Lagrange, Legendre, Germain, Mordell, Euler, Jacobi, Gauss, Möbius, Pell). Standard contents plus the Gaussian integers and the RSA-cypher system. A good choice for a textbook. SB

Number Theory, P. *Lectures on the Geometry of Numbers.* Carl Ludwig Siegel. Springer-Verlag, 1989, x + 160 pp, $45. [ISBN: 0-387-50629-2] A reprint of Siegel's 1945-46 New York University lecture notes as revised by K. Chandrasekharan. The fifteen lectures focus on three major themes: Minkowski's theorems, linear inequalities, and reduction theory. SG

Number Theory, P. *Irregularities of Partitions.* Ed: G. Halász, V.T. Sós. Algorithms & Combinatorics, V. 8. Springer-Verlag, 1989, 165 pp, $39.50 (P). [ISBN: 0-387-50582-2] The proceeding of a meeting held at Fertöd, Hungary from July 7-11, 1986. This meeting brought together number theorists and combinatorists concerned with problems involving the distribution of sequences of elements in certain collections of subsets. CEC

Number Theory, T(15-17: 1), S, P, L. *Factorization and Primality Testing.* David M. Bressoud. Undergrad. Texts in Math. Springer-Verlag, 1989, xiii + 237 pp, $45. [ISBN: 0-387-97040-1] Theoretical and computational considerations of modern methods in factorization and primality testing: Pollard's rho and $p - 1$ methods, the quadratic sieve, continued-fraction algorithms, and elliptic-curve techniques. The informal writing style and numerous examples and exercises make this book a good choice for anyone interested in factoring large numbers. BC

Number Theory, S(16-17), P*, L*. *The Book of Prime Number Records, Second Edition.* Paulo Ribenboim. Springer-Verlag, 1989, xxiii + 479 pp, $49.80. [ISBN: 0-387-97042-8] More than just a "Guinness Book" for mathematicians, this valiant attempt to keep track of advances on the prime number front presents a cogent and entertaining account of number theory and its recent (and some not-so-recent) developments (*First Edition*, TR, December 1988; Extended Review, August-September 1989). BC

Number Theory, P. *Lecture Notes in Mathematics-1395: Number Theory, Madras 1987.* Ed: K. Alladi. Springer-Verlag, 1989, vii + 234 pp, $20 (P). [ISBN: 0-387-51595-X] Nine papers directly related to the work of Ramanujan, primarily from a symposium commemorating his 100th birthday. GG

Number Theory, T(16-18: 1), S, P, L. *Binary Quadratic Forms: Classical Theory and Modern Computations.* Duncan A. Buell. Springer-Verlag, 1989, x + 247 pp, $35. [ISBN: 0-387-97037-1] The quadratic expression $ax^2 + bxy + cy^2$ is one of the most fruitful in all of mathematics. Gauss devoted most of his *Disquisitiones Arithmeticae* to unraveling its number-theoretic mysteries. This book is sort of an update of Gauss, using the modern language of abstract algebra (and also the "correct" definition of a binary quadratic form), but still emphasizing explicit computational methods. Try to imagine what Gauss would have done with a PC. BC

Number Theory, T(18: 1), S, P. *Holomorphic Hilbert Modular Forms.* Paul B. Garrett. Math. Ser. Wadsworth, 1990, xiii + 304 pp. [ISBN: 0-534-10344-8] An introduction to Hilbert modular forms, which generalize elliptic modular forms from the rational numbers to totally real fields, with implications for the analytic continuation and functional equations of Dirichlet series. Covers classical and adelic viewpoints, Eisenstein series, theta series, and the arithmetic of Hilbert modular forms. BC

Group Theory, P. *Lecture Notes in Mathematics-1379: Probability Measures on Groups IX.* Ed: H. Heyer. Springer-Verlag, 1989, viii + 437 pp, $37.50 (P). [ISBN: 0-387-51401-5] Thirty papers from a January 1988 conference held at Oberwolfach. GG

Group Theory, T(18: 1), S, P. *Topological Groups: Characters, Dualities, and Minimal Group Topologies.* Dikran N. Dikranjan, Ivan R. Prodanov, Luchezar N. Stoyanov. Pure & Appl. Math., V. 130. Marcel Dekker, 1990, x + 287 pp, $99.75. [ISBN: 0-8247-8047-7] An attempt to present a substantial theory for topological groups (at least for the Abelian case) without resorting to functional analysis. A good share of the presentation depends on work of Folner concerning existence of continuous characters. Applications include almost periodic functions, the Haar integral, duality, and minimal groups; both Abelian and non-Abelian. Chapter notes, exercises, bibliography, index. JS

Algebra, T(17: 1), S, P. *An Introduction to Noncommutative Noetherian Rings.* Math. Soc. Stud. Texts, V. 16. Cambridge University Pr, 1989, xvii + 303 pp, $49.50; $19.95 (P). [ISBN: 0-521-36086-2; 0-521-36925-8] Intended as a largely self-contained introduction to non-commutative Noetherian rings at the level of a second-year graduate student; discussion begins with the examples of polynomial identity rings, group algebras, differential operators, and enveloping algebras for Lie algebras. Theory includes prime ideals, modules, Goldie theorems, Krull dimension, division algebras. Exercises, notes, bibliography. JS

Algebra, P. *Commutative Algebra.* Eds: M. Hochster, C. Huneke, J.D. Sally. Math. Sci. Res. Inst., V. 15. Springer-Verlag, 1989, x + 516 pp, $49.80. [ISBN: 0-387-96990-X] Proceedings of a three-week program in Commutative Algebra held at the Mathematical Sciences Research Institute at Berkeley from June 15 to July 2, 1987. Contains 27 articles, some of which are expository and some which include current research. Should be of interest to both experts and non-experts in commutative algebra. RH

Algebra, T(18: 1), S, P. *Infinite Crossed Products.* Donald S. Passman. Pure & Appl. Math., V. 135. Academic Pr, 1989, xii + 468 pp, $84.50. [ISBN: 0-12-546390-1] A fairly self-contained book mainly concerned with recent developments in the theory of crossed products. Topics include Cohen-Montgomery duality, the symmetric Martindale ring of quotients, classification of prime and semiprime crossed products, the Galois theory of prime and semiprime rings, and the Grothendieck group of a Noetherian crossed product. Includes over 200 exercises and an excellent list of references. CEC

Calculus, S(13). *Introduction to Calculus and Analysis.* Richard Courant, Fritz John. Springer-Verlag, 1989. *Volume I,* xxiii + 661 pp, $45, [ISBN: 0-387-97151-3]; *Volume II,* xxiii + 954 pp, $59. [ISBN: 0-387-97152-1] First published in 1965 as an attempt to provide to the English speaking world a calculus text in the spirit of Courant's 1927 German text. Rigorous, yet moving quickly to the uses of calculus, the book is a source for some (not new) ideas one hears mentioned in calculus reform conferences: defining the integral as an area, relegating indeterminate forms and L'Hospital's rule to an appendix, minimal attention to the convergence of series of constants, more attention to complex analysis. Would not win a prize in the contest to write a book with the most exercises; neither is it a candidate for a "lean calculus" text. AWR

Calculus, T(16: 2). *Advanced Calculus, Revised Edition.* Lynn H. Loomis, Shlomo Sternberg. Jones & Bartlett, 1990, 580 pp, $40. [ISBN: 0-86720-122-3] A second course in advanced calculus, which presupposes a rigorous calculus course and a course in linear algebra. Provides a concise treatment of differentiation and integration in several variables. Mixed in are chapters on compactness and completeness, differential equations, multilinear functionals, manifolds, potential theory, and classical mechanics. Numerous exercises, including some of a computational nature. (*First Edition,* TR, January 1968; Extended Reviews, April 1970 and October 1970.) GG

Real Analysis, T(18: 1, 2), S, P. *On L^1-approximation.* Allan M. Pinkus. Tracts in Math., V. 93. Cambridge University Pr, 1989, x + 239 pp, $44.50. [ISBN: 0-521-36650-X] A clearly written, friendly, and enthusiastic introduction to the qualitative theory of best approximation, in the L^1 sense, from finite dimensional subspaces. Problems are initially posed quite generally, but much of the development concerns the special case approximation of continuous functions on compact sets endowed with tractable measures. Includes a significant number of concrete examples and exercises; some of the latter develop material used elsewhere in the text. PZ

Partial Differential Equations, P. *Lecture Notes*

in Mathematics-1390: Stochastic Partial Differential Equations and Applications II. Eds: G. Da Prato, L. Tubaro. Springer-Verlag, 1989, vi + 258 pp, $24.30 (P). [ISBN: 0-387-51510-0] Proceedings from the second conference on this topic held in Italy in 1988. (The proceedings of the first conference are in *LNM-1236* of this series.) Twenty-two short papers. JO

Partial Differential Equations, P. *Multilevel Adaptive Methods for Partial Differential Equations.* Stephen F. McCormick. Frontiers in Appl. Math., V. 6. SIAM, 1989, ix + 162 pp, $24.50 (P). [ISBN: 0-89871-247-5] Methods designed for discretization and solution of partial differential equations. Includes the finite volume method and multigrid methods (common discretization methods for computational fluid dynamics applications). MLR

Partial Differential Equations, P. *Nonlinear Boundary Value Problems in Science and Engineering.* C. Rogers, W.F. Ames. Math. in Sci. & Eng., V. 183. Academic Pr, 1989, xiii + 417 pp, $69.96. [ISBN: 0-12-593110-7] "In this text ... an attempt has been made to survey those nonlinear boundary value problems amenable to exact solutions." This book is stuffed full of exact solutions obtained by symmetry, series, and constitutive law methods. Of particular interest is the appendix with full symmetry groups for forty differential equations. JO

Numerical Analysis, P. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics.* Roland Glowinski, Patrick Le Tallec. Stud. in Appl. Math., V. 9. SIAM, 1989, x + 295 pp, $44.50. [ISBN: 0-89871-230-0] Augmented Lagrangian methods generalize Lagrange multipliers and lead to optimization algorithms with improved convergence properties. This book applies numerical algorithms based on augmented Lagrangians to problems in elasticity and fluid mechanics. JO

Numerical Analysis, T(18: 2), P. *Approximation by Spline Functions.* Günther Nürnberger. Springer-Verlag, 1989, xi + 243 pp, $39.50. [ISBN: 0-387-51618-2] Studies in a unified approach to the basic theoretical and numerical aspects of interpolation and best approximation by polynomial splines in one variable. Prerequisites are basic knowledge of analysis and linear algebra. LC

Numerical Analysis, P. *Lecture Notes in Mathematics-1349: Deterministic and Stochastic Error Bounds in Numerical Analysis.* Erich Novak. Springer-Verlag, 1988, 113 pp, $13.10 (P). [ISBN: 0-387-50368-4] Covers deterministic error bounds, error bounds for Monte Carlo methods, and average error bounds. LC

Numerical Analysis, P. *Lecture Notes in Mathematics-1397: Numerical Analysis and Parallel Processing.* Ed: P.R. Turner. Springer-Verlag, 1989, vi + 264 pp, $24.30 (P). [ISBN: 0-387-51645-X] Notes for five of the eight series of lectures presented at the Third Science and Engineering Research Council Numerical Analysis Summer School held July 12-31, 1987 at the University of Lancaster. Authors are L.C.W. Dixon, J.A. George, C.W. Clenshaw, F.W.J. Olver and P.R. Turner, A. Feldstein and R.H. Goodman, J.R. Whiteman and G. Goodsell. DFA

Numerical Analysis, S(18), P. *Lecture Notes in Physics-318: An Introduction to the Numerical Analysis of Spectral Methods.* Bertrand Mercier. Springer-Verlag, 1989, v + 154 pp, $23.30. [ISBN: 0-387-51106-7] Begins with Fourier series, continues with polynomial expansions (in particular Chebyshev expansions), interpolation operators, and Fast Fourier Transforms. Includes a rigorous treatment of one-dimensional advection and diffusion equations. Terse, with few examples and no exercises. MPR

Numerical Analysis, P. *Adaptive Methods for Partial Differential Equations.* Eds: Joseph E. Flaherty, *et al.* SIAM, 1989, 265 pp, $31.50 (P). [ISBN: 0-89871-242-4] "The essential part of any adaptive approach is the *a posteriori* determination of the error. If the error is not acceptable, then an adaptation of the computational process is made." Sixteen papers from the Workshop on Adaptive Computational Methods for Partial Differential Equations held at Rensselaer Polytechnic Institute in 1988. JO

Functional Analysis, T(17-18: 3), L**.** *Functional Equations in Several Variables with Applications to Mathematics, Information Theory and to the Natural and Social Sciences.* J. Aczél, J. Dhombres. Ency. of Math. & Its Applic., V. 31. Cambridge U Pr, 1989, xiv + 462 pp, $89.50. [ISBN: 0-521-35276-2] The emphasis is on applications of the theory. It is organized so that most of the chapters are relatively independent of each other. The prerequisites are calculus through basic Lebesgue integration and linear algebra. Over 400 exercises, some of which require outside references to solve. Also includes 25 pages of historical notes and an extensive bibliography (ordered chronologically). MPR

Functional Analysis, P. *Lectures on the Ekeland Variational Principle with Applications and Detours.* D.G. de Figueiredo. Springer-Verlag, 1989, 96 pp, $11.90 (P). [ISBN: 0-387-51179-2] A consolidation of results on the subject, with applications to semilinear elliptic differential equations and the geometry of Banach spaces. JO

Functional Analysis, P. *Nonlinear Functional Analysis.* Ed: P.S. Milojević. Lect. Notes in Pure & Appl. Math., V. 121. Marcel Dekker, 1990, vii + 264 pp, $89.75 (P). [ISBN: 0-8247-8255-0] Seven papers presented at the Special Session on Nonlinear Functional Analysis held at the New Jersey Institute of Technology, April 1987. KS

Functional Analysis, T(18), P. *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.* Wiliam P. Ziemer. Grad. Texts in Math., V. 120. Springer-Verlag, 1989, xvi + 308 pp, $49. [ISBN: 0-387-97017-7] A weakly differentiable function is integrable and has partial derivatives in the sense of distributions that are L^p functions (Sobolev functions) or (signed) measures with finite total variation. This book analyzes the pointwise behavior of these functions. Historical notes include references to important and relatively new

results. MLR

Functional Analysis, T(16-17: 2), L. *Introductory Functional Analysis with Applications.* Erwin Kreyszig. Robert E Krieger, 1989, xiv + 688 pp, $56. [ISBN: 0-89464-332-0] Unaltered reprint of a 1978 Wiley publication (TR, March 1978). LAS

Analysis, P, L*. *Analysis I: Integral Representations and Asymptotic Methods.* Ed: R.V. Gamkrelidze. Encyclop. of Math. Sci., V. 13. Springer-Verlag, 1989, 238 pp, $59.50. [ISBN: 0-387-17008-1] A smoothly moving account of developments starting with Newton and Euler, this can be recommended to someone wanting to know what ideas led up to, say, attention to divergent integrals. Could serve as either an introduction or a capstone to a course in analysis. AWR

Analysis, P. *Lecture Notes in Mathematics-1384: Harmonic Analysis and Partial Differential Equations.* Ed: J. García-Cuerva. Springer-Verlag, 1989, vii + 213 pp, $20 (P). [ISBN: 0-387-51460-0] Proceedings of a conference held in Spain in June 1987. Four survey lectures and ten more specialized papers, one in French. JO

Analysis, P. *Theory of Reproducing Kernels and its Applications.* Saburou Saitoh. Pitman Res. Notes in Math. Ser., V. 189. Longman Scientific & Technical (US Distr: Wiley), 1988, 157 pp, $57.95 (P). [ISBN: 0-582-03564-3] Aims to introduce the theory of reproducing kernels from two viewpoints: the general theory as established by N. Aronszajn, and the theories of complex analysis of one variable as developed by S. Bergman, M. Schiffer, P. Garabedian, Z. Nehari, D. Hejhal, etc. LC

Analysis, P. *Lecture Notes in Mathematics-1393: Séminaire de Théorie du Potentiel Paris, No. 9.* Eds: N. Bouleau, *et al.* Springer-Verlag, 1989, vi + 265 pp, $24.30 (P). [ISBN: 0-387-51592-5] This volume emphasizes the connections with other branches of analysis. JAS

Analysis, T(18: 2), P. *Asymptotic Approximations of Integrals.* R. Wong. Comput. Sci. & Scientific Comput. Academic Pr, 1989, xiii + 543 pp, $69.95. [ISBN: 0-12-762535-6] Studies methods used in asymptotic approximation of integrals, including Mellin transform technique for multiple integrals, summability method, distributional approach, logarithmic singularities, uniform asymptotic expansions via a rational transformation, and double integrals with a curve of stationary points. Prerequisites are advanced calculus and complex variables. LC

Analysis, T(13-14: 1, 2), L.** *A Course in Mathematics For Students of Physics 1.* Paul Bamberg, Shlomo Sternberg. Cambridge U Pr, 1988, xvii + 405 pp, $49.50. [ISBN: 0-521-25017-X] Developed at Harvard over the last ten years for students with a solid background in one-variable calculus. If you do not have a course for which this book is suitable, you might want to introduce one. Best for students concurrently studying physics. Covers the theory and *physical* applications of linear algebra and several variable calculus (in particular, the exterior calculus). Written in "spiral method." Chapter summaries and exercises. JDEK

Analysis, T(17-18), S, P. *An Introduction to Harmonic Analysis on Semisimple Lie Groups.* V.S. Varadarajan. Stud. in Adv. Math., V. 16. Cambridge U Pr, 1989, x + 316 pp, $69.50. [ISBN: 0-521-34156-6] Presents the representation theory of semisimple groups via special examples. Throughout, expository writing traces the historical development of the subject and makes this a very readable account. As a self-contained introduction with problems at the end of each chapter, this book can serve as a text or as a researcher's guide to this central part of mathematics. KS

Algebraic Geometry, P. *The Curves Seminar at Queen's, Volume VI.* Anthony V. Geramita. Papers in Pure & Appl. Math., No. 83. Queen's U, 1989, 319 pp, (P). Contains three substantial expository articles on the non-singular cubic surface in P^3 (A. Geramita), fat points (A. Gimigliano), and liaison (J. Migliore), along with nine research articles. SG

Algebraic Geometry, P. *Lecture Notes in Mathematics-1392: Real Algebraic Surfaces.* Robert Silhol. Springer-Verlag, 1989, x + 215 pp, $20 (P). [ISBN: 0-387-51563-1] An investigation of the general problem of classifying real algebraic surfaces and the specific problem of determining the topology of the real part of a real algebraic surface. SG

Algebraic Geometry, P. *Group Actions and Invariant Theory.* Eds: A. Białynicki-Birula, *et al.* Conf. Proc., V. 10. AMS, 1989, x + 228 pp, (P). [ISBN: 0-8218-6015-1] Twenty-two papers on groups acting on algebraic spaces. The papers were presented at a conference in Montreal in 1988. SG

Algebraic Geometry, P. *2-Knots and their Groups.* Jonathan Hillman. Australian Math. Soc. Lect. Ser., V. 5. Cambridge U Pr, 1989, x + 164 pp, $19.95 (P). [ISBN: 0-521-37812-5] An n-knot is an embedding $K : S^n \to S^{n+2}$. (For $n = 1$ we have the "classical knots.") The major result contained herein is that the 4-manifold obtained by surgery on a 2-knot is often aspherical. This is in contrast to the fact that the exterior of a nontrivial n-knot ($n > 1$) is never aspherical. Also included is a discussion of how to recover 2-knots from their group invariants. It is shown that this is possible if the group is torsion free and polycyclic. MPR

Differential Geometry, P. *Cosmology in* $(2 + 1)$*-Dimensions, Cyclic Models, and Deformations of* $M_{2,1}$. Victor Guillemin. Annals of Math. Stud., No. 121. Princeton U Pr, 1989, v + 228 pp, $50; $15 (P). [ISBN: 0-691-08513-7] An exposition of Zollfrei manifolds in dimension three with the aim of providing a low-dimensional framework for various more general questions in cosmology. JAS

Differential Geometry, T*, S(18), P. *The Method of Equivalence and Its Applications.* Robert B. Gardner. CBMS-NSF Reg. Conf. Ser. in Appl. Math., V. 58. SIAM, 1989, vii + 127 pp, $21.75 (P). [ISBN: 0-89871-240-8] The problem of equivalence is to characterize conditions under which sets of ge-

ometric objects are equivalent, i.e., there exist diffeomorphisms mapping one set onto the other. The author restricts himself to cases where the diffeomorphisms may be described as solutions to first order systems of differential equations. In particular, the focus is on the case where sets of geometric objects are control systems and the equivalences are feedback transformations. This appears to be nicely written (in the form of ten lectures) with many examples included to illustrate ideas. No exercises. MPR

Differential Geometry, S(18), L. *The Theory and Applications of Harmonic Integrals.* W.V.D. Hodge. Math. Library. Cambridge University Pr, 1989, xiii + 284 pp, $22.95 (P). [ISBN: 0-521-35881-7] Paperback edition of this classic is enhanced by a Foreword by Sir Michael Atiyah in which he sets Hodge's work in its historical context and connects it to subsequent developments. Geometry of manifolds and applications to algebraic varieties and continuous groups. Takes readers back to the origins of the subject. Not to be read for currency. JDEK

Differential Geometry, T(17-18: 2, 3). *Differential Geometry, Gauge Theories, and Gravity.* M. Göckeler, T. Schücker. Mono. on Math. Physics. Cambridge University Pr, 1989, xii + 230 pp, $19.95 (P). [ISBN: 0-521-37821-4] Intended as a graduate text in theoretical physics, this is a self-contained but terse presentation of the mathematics required for the modern treatment of gravity and general relativity. Although chapters on differential forms, manifolds, Lie groups, and fibre bundles limit the formal prerequisites to linear algebra and elementary real analysis, undergraduates should approach this book with extreme caution. Sections on Maxwell's equations, Einstein-Cartan theory, Yang-Mills theory, and monopoles and instantons provide the physical payoff for the mathematical effort. (1987 hardcover edition, TR, December 1988.) JO

Differential Geometry, S(17-18), P, L. *Lecture Notes in Mathematics-1000: Differential Geometry in the Large.* Heinz Hopf. Springer-Verlag, 1989, vii + 184 pp, $16.20 (P). [ISBN: 0-387-51497-X] Corrected reprint of the *LNM-1000* anniversary volume (TR, April 1984; Extended Review, January 1986), containing notes from 1946 and 1956 lectures by Hopf, prepared by Peter Lax and John Gray. New Preface contains pointers to recent work by Henry Wente and others on a new class of surfaces in E^3 of constant mean curvature, resolving questions posed by Hopf in these notes. LAS

Algebraic Topology, P. *Lecture Notes in Mathematics-1370: Algebraic Topology.* Ed: G. Carlsson, *et al.* Springer-Verlag, 1989, ix + 456 pp, $47 (P). [ISBN: 0-387-51118-0] Proceedings of an international conference held in Arcata, California, July 27-August 2, 1986. Special attention is given to the silver anniversary of *Topology* and the work of E.H. Brown. JAS

Topology, P. *Lecture Notes in Mathematics-1374: The Topology of 4-Manifolds.* Robion C. Kirby. Springer-Verlag, 1989, vi + 108 pp, $13.50 (P). [ISBN: 0-387-51148-2] Exposition of a number of improved proofs and results concerning cobordism groups of 4-manifolds with special attention given to Rohlin's theorem: $\Omega_4^{SO} = Z$. JAS

Topology, S*, L*. *Mathematics and Knots.* Ronnie Brown, *et al.* The Mathematics and Knots Exhibition Group (U. of Wales, School of Mathematics, Dean Street, Bangor, Gwynedd LL57 1UT, Wales, UK), 1989, 27 pp, (P). [ISBN: 0-9514947-16] Text with illustrations of a museum exhibit on classification of knots prepared at the University of Wales. Illustrates crossing numbers, unknotting numbers, bridge numbers, coloring, links, mirror images, arithmetic of knots, prime knots, and diverse applications (fluid flow, string theorem, DNA). Excellent visual introduction. LAS

Topology, T(17-18: 1, 2). *Fibrewise Topology.* I.M. James. Tracts in Math., V. 91. Cambridge U Pr, 1989, x + 198 pp, $49.50. [ISBN: 0-521-36090-0] An exposition of general topology (through homotopy) taking the fundamental objects to be continuous functions from spaces X to a fixed base space B. This provides a framework to bring ideas from fiber spaces to general topology. Exercises, references, and index. JAS

Operations Research, S(11-13). *Queues: Will This Wait Never End!* Clifford Sloyer, *et al.* Contemp. Appl. Math. Janson, 1987, vi + 41 pp, $7.95 (P). [ISBN: 0-939765-08-X] A high-school level workbook approach to queues based on a planned sequence of 77 questions (with answers at the back) interspersed with minimal exposition. Begins with constant arrival rates, and ends with simulations of variable arrivals and service times. LAS

Optimization, T(14-15: 1), L. *Linear Programming and Its Applications.* James K. Strayer. Undergrad. Texts in Math. Springer-Verlag, 1989, xi + 265 pp, $47. [ISBN: 0-387-96930-6] Covers solving linear programming problems geometrically and by the simplex algorithm using the Tucker tableau. Also includes duality theory and problems that are not immediately solvable by the simplex algorithm. Includes applications to game theory, transportation and assignment problems, and network-flow problems. Has many examples and exercises. Prerequisites are a knowledge of linear equations, including the graphing of lines and planes, and the solution of systems of simultaneous linear equations without matrices. RH

Optimization, P. *Finite Element Approximation for Optimal Shape Design: Theory and Applications.* J. Haslinger, P. Neittaanmäki. Wiley, 1988, xii + 335 pp, $79.95. [ISBN: 0-471-92079-7] Investigates "optimal design problems via optimal control theory where the states of systems are governed by variational inequalities." LC

Dynamical Systems, S(18), P. *Vacillation and Predictability Properties of Low-Order Atmospheric Spectral Models.* H.E. de Swart. CWI Tract, V. 60. Mathematisch Centrum, 1989, 121 pp, Dfl. 19 (P). [ISBN: 90-6196-368-0] A low-order spectral model

of the atmosphere is obtained when the infinite number of nonlinear coupled ordinary differential equations arising from an eigenfunction expansion are truncated to a finite system. One criterion for the truncation is that the resulting model should yield trajectories tending towards strange attractors. This tract discusses models truncated to 3, 6, and 10 components. One conclusion: "The atmosphere possesses an intrinsic finite range of predictability of about two weeks." MPR

Dynamical Systems, S(18), P. *Dynamical Systems II.* Ed: Ya. G. Sinai. Encyclop. of Math. Sci., V. 2. Springer-Verlag, 1989, viii + 281 pp, $59.95. [ISBN: 0-387-17001-4] More on dynamical systems and differential equations. From the Russian school. This volume contains three articles: General Ergodic Theory of Groups of Measure Preserving Transformations; Ergodic Theory of Smooth Dynamical Systems; and Dynamical Systems of Statistical Mechanics and Kinetic Equations. MPR

Dynamical Systems, T, S*(16-18), P, L*. *Perspectives of Nonlinear Dynamics, Volume 1.* E. Atlee Jackson. Cambridge U Pr, 1989, xix + 495 pp, $54.50. [ISBN: 0-521-34504-9] An eclectic collage of introductory examples, results, and applications in nonlinear systems organized roughly in chronological terms. "In the beginning... there was Poincaré." Touches on an enormous variety of stimulating ideas, including catastrophe theory, fractals, Lyapunov exponents, deterministic chaos, forced oscillators, cellular automata. Illustrated with diverse applications to dozens of different fields. Several appendices provide additional mathematical depth. Massive 1600 item bibliography. Superbly crafted for its intended purpose—to stimulate imagination; useful as a systematic text. LAS

Dynamical Systems, S(18), P. *Lecture Notes in Control and Information Sciences-127: Deterministic Identification of Dynamical Systems.* C. Heij. Springer-Verlag, 1989, 292 pp, $41.60 (P). [ISBN: 0-387-51323-X] The aim of this monograph is to show how a model, especially a dynamical system, may be determined from a set of data, especially a time series. In this case, the models are deterministic because there is no underlying probabilistic or stochastic assumptions. Uncertainties and errors which arise are due to the complexities of reality and the inherent shortcomings of the model. MPR

Systems Theory, P, L. *Lecture Notes in Control and Information Sciences-134: Auziliary Signal Design in Fault Detection and Diagnosis.* X.J. Zhang. Springer-Verlag, 1989, xii + 213 pp, $30.30 (P). [ISBN: 0-387-51559-3] To avoid catastrophies such as the Three Mile Island Accident, systems to detect and diagnose faults must be efficient yet not overly complex. The author develops one such system and applies it to the anhydrous caustic soda plant. KS

Systems Theory, P. *The Volterra and Wiener Theories of Nonlinear Systems.* Martin Schetzen. Robert E Krieger, 1989, xviii + 573 pp, $55.50. [ISBN: 0-89464-356-8] Reprint of a 1980 volume, up-dated with two papers added in the appendices. In one, Volterra theory is used to develop a new theory for analysis of nonlinear networks. The other new paper introduces the concept of differential models and develops it using Wiener theory. KS

Probability, T*(1-3), S, P, L. *Introduction to Probability Models, Fourth Edition.* Sheldon M. Ross. Academic Pr, 1989, xiv + 544 pp, $44.50. [ISBN: 0-12-598464-2] Differs from *Third Edition* (TR, December 1985) by addition of material on Markov chains, minimal theory, and simulation. Postcalculus introduction to probability theory, stochastic processes, queueing theory, and simulation. (*First Edition,* TR, March 1973; Extended Review, November 1975; *Second Edition,* TR, January 1982.) TH

Probability, P. *Central Limit Theorems for Generalized Multilinear Forms.* P. de Jong. CWI Tract, V. 61. Mathematisch Centrum, 1989, 84 pp, Dfl. 14.10 (P). [ISBN: 90-6196-369-9] Theoretical monograph presenting "some central limit theorems for homogeneous multilinear forms and for generalizations of multilinear forms." RSK

Probability, P. *Lecture Notes in Mathematics-1391: Probability Theory on Vector Spaces IV.* Eds: S. Cambanis, A. Weron. Springer-Verlag, 1989, viii + 424 pp, $37.50 (P). [ISBN: 0-387-51548-8] 34 papers from a June 1987 conference in Łańcut, Poland on probability measures on algebraic and topological structures, limit theorems on vector spaces, infinitely divisible measures, and stochastic integrals. TH

Stochastic Processes, P. *Stochastic Differential Equations and Diffusion Processes, Second Edition.* Nobuyuki Ikeda, Shinzo Watanabe. Math. Lib., V. 24. North-Holland (US Distr: Elsevier Science), 1989, xvi + 555 pp, $147.25. [ISBN: 0-444-873783] Changes from the *First Edition* (TR, April 1982) include Brownian motion excursion and Malliavin calculus sections, and added sections on complex martingales and Kähler diffusions. Terse, rigorous presentation. TH

Stochastic Processes, P. *Lecture Notes in Mathematics-1372: Séminaire de Probabilités XXIII.* Eds: J. Azéma, P.A. Meyer, M. Yor. Springer-Verlag, 1989, iv + 582 pp, $49.50 (P). [ISBN: 0-387-51191-1] 27 French and 18 English papers from a seminar in Strasbourg. Diffusion processes, martingales, Brownian motion. TH

Elementary Statistics, T(13-14: 2). *Statistical Techniques in Business and Economics, Seventh Edition.* Robert D. Mason, Douglas A. Lind. Richard D Irwin, 1990, xvii + 910 pp, $43.95. [ISBN: 0-256-07696-0] Added features in the new edition include stem-and-leaf displays, the finite population correction factor, two-factor ANOVA, sign tests, acceptance sampling, and MINITAB output. (*Fourth Edition,* TR, November 1978; *Fifth Edition,* TR, November 1982; *Sixth Edition,* TR, August-September 1986.) RWJ

Elementary Statistics, T(14: 1), S. *Statistics for Biologists, Third Edition.* R.C. Campbell. Cam-

bridge U Pr, 1989, xviii + 446 pp, $59.50. [ISBN: 0-521-36095-1] Revision of the author's 1974 *Second Edition* (TR, January 1975). Main change is the addition of computer analyses of a selection of the examples, using several different statistical packages. RSK

Computational Statistics, P? *Spatial Regression Analysis on the PC: Spatial Statistics Using Minitab.* Daniel A. Griffith. Discussion Paper No. 1. Institute of Mathematical Geography (2790 Briarcliff, Ann Arbor, MI 48105), 1989, vi + 84 pp, $12.95 (P). Gives Minitab code to implement spatial statistics estimates (e.g., Moran coefficient), but gives no indication as to what is being estimated. TH

Computational Statistics, P. *A Specification System for Statistical Software.* V.J. de Jong. CWI Tract, V. 62. Mathematisch Centrum, 1989, 250 pp, Dfl. 37.90 (P). [ISBN: 90-6196-370-2] Describes CONDUCTOR, a very high-level (fourth-generation) language for statistics, designed to be more flexible than statistical packages, more efficient than S or SAS/IML, and have lower life-cycle costs than low-level language programs. Concepts, examples (the bootstrap), and a formal description of the language. TH

Statistics, T?(13-14), S*. *Modern Elementary Probability and Statistics with Statistical Programming in SAS, MINITAB & BMDP.* Edward J. Dudewicz, Pinyuen Chen, Baldeo K. Taneja. American Sciences Pr, 1989, viii + 375 pp, $34.50 (P). [ISBN: 0-935950-19-2] Rich set of realistic problems; book written for mathematically unsophisticated. Better editing and printing would significantly enhance the value. Omits fundamental concepts such as difference between discrete and continuous random variables; introduces material best left for a different course, e.g., Bayes theorem. Not a good introductory book for someone planning to take another course in statistics. WE

Statistics, S(15-18), L. *Lecture Notes in Statistics-56: The Analysis of Categorical Data Using GLIM.* James K. Lindsey. Springer-Verlag, 1989, v + 168 pp, $19.50 (P). [ISBN: 0-387-97029-0] GLIM 3.77 update 2 is applied to a variety of examples in the social sciences for which log-linear/logistic models are used. Assumes previous knowledge of log-linear/logistic models. Appendices include annotated GLIM programs and macros. RWJ

Statistics, T(18: 1), P*, L. *Asymptotic Techniques for Use in Statistics.* O.E. Barndorff-Nielsen, D.R. Cox. Mono. on Stat. & Appl. Prob. Chapman & Hall, 1989, x + 252 pp, $49.95. [ISBN: 0-412-31400-2] Survey of asymptotic approximation methods for the distribution of (functions of) means of random variables, e.g., delta method. Uni- and multi-variate; conversational tone; few proofs. TH

Statistics, S(16-18). *A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models.* Barbara M. Byrne. Springer-Verlag, 1989, xii + 184 pp, $49. [ISBN: 0-387-96972-1] Non-technical introduction to the use of the LISREL (linear structural relations) VI program by Jöreskog and Sörbom (1985). Sample program input and output displayed for "self-concept" investigations in psychology. RWJ

Statistics, T(17-18: 1), L. *The Statistical Analysis of Discrete Data.* Thomas J. Santner, Diane E. Duffy. Texts in Stat. Springer-Verlag, 1989, xii + 367 pp, $45. [ISBN: 0-387-97018-5] Assumes some background in decision theory and linear model theory for continuous data. Covers point and interval estimates of binomial, multinomial, and Poisson parameters, loglinear models, cross-classified data, logistic regression, two by T and two by two by S tables. Includes recent advances in Bayes and shrinkage estimation, graphical models for contingency tables, and diagnostic tools for loglinear models and logistic regression. RWJ

Statistics, P. *Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin.* Eds: Leon Jay Gleser, *et al.* Springer-Verlag, 1989, x + 505 pp, $69. [ISBN: 0-387-97076-2] Collection of 32 original essays by former students, collaborators, and colleagues of Olkin, divided into four areas: probability inequalities and characterizations; multivariate analysis and association; linear and nonlinear models, ranking and selection, design; and approaches to inference. Also includes an interview with Olkin, a bibliography, and a short biography. RSK

Statistics, P. *Probability, Statistics, and Mathematics: Papers in Honor of Samuel Karlin.* Eds: T.W. Anderson, Krishna B. Athreya, Donald L. Iglehart. Academic Pr, 1989, xl + 371 pp, $39.95. [ISBN: 0-12-058470-0] A collection of 24 papers written by mathematicians, probabilists, statisticians, and economists on the occasion of Samuel Karlin's 65th birthday. Includes a list of Karlin's publications and students. RWJ

Statistics, T(17: 2). *Fundamentals of Mathematical Statistics.* H.T. Nguyen, G.S. Rogers. Texts in Stat. Springer-Verlag, 1989, $39.80 each. *Volume I: Probability for Statistics,* x + 432 pp [ISBN: 0-387-97014-2]; *Volume II: Statistical Inference,* xi + 422 pp. [ISBN: 0-387-97020-7] Classical text, with the material divided into lessons, each appropriate for a single class period. *Volume I* is divided into three sections: elementary probability and statistics; probability and expectation; and limiting distributions. *Volume II* contains four sections: sampling and distributions; statistical estimation; testing hypotheses; and special topics. Exercises are integrated into the discussions. RSK

Statistics, T(15: 1, 2). *Probability and Statistics: Theory and Applications.* Gunnar Blom. Texts in Stat. Springer-Verlag, 1989, xi + 356 pp, $59. [ISBN: 0-387-96852-0] Modified translation of a Swedish text. Interesting alternative to the usual calculus-based texts. Emphasizes practical applications while holding mathematical arguments to a minimum. Includes chapters on descriptive statistics and planning statistical investigations, but omits topics such as analysis of variance, moment generat-

ing functions, and the F-distribution. RSK

Statistics, S(14-15). *Concise Statistics.* M.G. Godfrey, E.M. Roebuck, A.J. Sherlock. Edward Arnold, 1988, vi + 402 pp, $19.95 (P). [ISBN: 0-7131-3591-3] Study guide for statistics for advanced level mathematical syllabuses in England. Contains numerous exercises with fairly complete answers. RSK

Statistics, S(18). *Multivariate Statistics and Probability: Essays in Memory of Paruchuri R. Krishnaiah.* Eds: C.R. Rao, M.M. Rao. Academic Pr, 1989, xiii + 567 pp, $59.95. [ISBN: 0-12-580205-6] A collection of 35 papers covering the main areas of multivariate statistical theory and its applications, along with some probability and stochastic analysis. The papers discuss finite sampling and asymptotic results, including decision theory, Bayesian analysis, maximum likelihood estimation, regression, and time series. Most results theoretical, although some practical applications are included. Of interest to researchers and graduate students in theoretical and applied statistics, multivariate analysis, and random processes. Papers from 61 authors from several countries. JJ

Statistics, P, L. *Proceedings of the Thirty-Fourth Conference on the Design of Experiments.* US Army Research Office (POB 12211, Research Triangle Park, NC), 1988, xi + 367 pp, (P). Twenty papers from an October 1988 conference held at New Mexico State University. LAS

Computer Literacy, S(13), L. *A Glossary of Computing Terms, An Introduction, Sixth Edition.* British Computer Society. Cambridge U Pr, 1989, xvi + 92 pp, $5.95 (P). [ISBN: 0-521-37828-1] Over 800 concise definitions, divided into 15 categories of origin (applications, data representation and structure, machine architecture, systems software, etc.). Suitable as a reference for the lay computer-user. A thin volume, but note its price! DFA

Programming, S. *An Easy Course in Using the HP-42S.* Dan Coffin, Chris Coffin. Grapevine Pub, 1989, 379 pp, $22 (P). [ISBN: 0-931011-26-4] Detailed handbook for using and programming the HP-42S, a hand-calculator with built-in capabilities for matrix and statistical calculations, solving equations and numerical integration. Includes alphabetic registers for naming variables and programs, branching, conditional tests, loops, indirect addressing, and reasonable editing capabilities for stored programs. LAS

Algorithms, P. *Lecture Notes in Mathematics-1387: Iterative Methods for Simultaneous Inclusion of Polynomial Zeros.* Miodrag Petković. Springer-Verlag, 1989, x + 263 pp, $24.30 (P). [ISBN: 0-387-51485-6] Galois' famous theorem states that explicit formulas determining zeros exist only for polynomials of degree less than five. Numerical methods must be used for higher degree polynomials. This book collects algorithms of the past fifteen years. Much of the work is in the spirit of the so-called "interval arithmetic" pioneered by Henrici, Gargantini, Alefeld, and Herzberger. MPR

Computer Systems, P. *The Society of Text: Hypertext, Hypermedia, and the Social Construction of Information.* Ed: Edward Barrett. Inform. Systems. MIT Pr, 1989, xix + 459 pp, $37.50. [ISBN: 0-262-02291-5] A sequel to *Text, ConText, and HyperText* (TR, June-July 1989) containing 22 new essays that together "assert a syntheses of three related areas of activity within the computer industry: writing, management, and engineering." These authors view computer-based information (including hypertext) as a form of social construction. Based in part on papers presented at an annual MIT summer conference "Writing for the Computer Industry." LAS

Computer Graphics, P, L. *Mathematical Methods in Computer Aided Geometric Design.* Ed: Tom Lyche, Larry L. Schumaker. Academic Pr, 1989, xv + 611 pp, $49.95. [ISBN: 0-12-460515-X] Forty-five papers from an international conference on mathematical methods in CAGD. Topics include varieties of splines (discrete D^m, multivariate, box, beta, parametric, etc.), Bezier curves, interpolation schemes, etc. BC

Artificial Intelligence, P. *Ontic: A Knowledge Representation System for Mathematics.* David A. McAllester. MIT Pr, 1989, xiv + 158 pp, $27.50. [ISBN: 0-262-13235-4] Ontic is a software system that automatically verifies formal mathematical theorems. It can both verify existing proofs as a check of their validity as well as try to discover proofs of currently unproven mathematical conjectures. It uses many of the new ideas of artificial intelligence to perform its proofs, including object-oriented inference and forward-chaining logic. The goal of the program is to demonstrate the validity of these new ideas by placing them in a working system for automated theorem proving. GMS

Computer Science, P. *Lecture Notes in Control and Information Sciences-130: Advances in Computing and Control.* Eds: W.A. Porter, S.C. Kak, J.L. Aravena. Springer-Verlag, 1989, vi + 367 pp, $47 (P). [ISBN: 0-387-51425-2] A selection of papers from an international conference on Advances in Communication and Control Systems held in Baton Rouge, Louisiana, October 19-21, 1988. A second volume, *Advances in Communications and Signal Processing,* contains other papers from this conference. JAS

Computer Science, P. *Resolution of Equations in Algebraic Structures, Volume 1: Algebraic Techniques.* Eds: Hassan Aït-Kaci, Maurice Nivat. Academic Pr, 1989, xxi + 452 pp, $64.95. [ISBN: 0-12-046370-9] Papers describing results, methods, algorithms and the state-of-the-art in resolution of equations in algebraic structures relevant to symbolic computation and the foundations of programming. By participants in the May 1987 colloquium in Lakeway, Texas. Contains practical applications as well as theoretical methods. A second collection from the same conference will concern rewriting techniques. DFA

Computer Science, P. *Categories in Computer*

Science and Logic. Ed: John W. Gray, Andre Scedrov. Contemp. Math., V. 92. AMS, 1989, x + 382 pp, $42 (P). [ISBN: 0-8218-5100-4] Papers from a conference held June 14-20, 1987 at the University of Colorado, Boulder. JAS

Applications, P*. *Theta Functions, Bowdoin 1987.* Eds: Leon Ehrenpreis, Robert C. Gunning. Proc. of Symposia in Pure Math., V. 49, Part 1 & 2. AMS, 1989 [ISBN: 0-8218-1485-0]. Part 1, x + 718 pp, $99; Part 2, x + 366 pp, $58. Very interesting collection of papers illustrating the wide range of applications of theta functions. Areas covered include infinite analysis, statistical mechanics, Kac-Moody algebras, algebraic geometry, and number theory. The level of the papers range from the very technical to survey level. MPR

Applications (Biological Science), S*(15), P, L*. *Biometrical Interpretation: Making Sense of Statistics in Biology, Second Edition.* Neil Gilbert. Oxford U Pr, 1989, viii + 146 pp, $21.95 (P). [ISBN: 0-19-854250-X] This edition reflects availability of computers for analysis; quantitative genetics and population biology chapters are rewritten. Conversational tone, emphasis on how to choose right analysis and make sense of the answer. Good supplemental reading for biologists who know some statistics. TH

Applications (Biological Science), T(17-18: 1, 2), S, P*, L*. *Mathematical Biology.* J.D. Murray. Springer-Verlag, 1989, xiv + 767 pp, $59. [ISBN: 0-387-19460-6] An excellent demonstration of how to use mathematics to model real biological phenomena. Treats population and epidemic models, biological oscillators and waves, diffusion mechanisms, and pattern formation (e.g., how the leopard got its spots—see page 436). Worth browsing through just for the non-mathematical discussions of butterfly wings, the Bubonic Plague, hallucinations, and the Lake Victoria Nile perch catastrophe. Well worth the price. BC

Applications (Communication Theory), T(17) S, P. *Digital Signal Processing: Applications to Communications and Algebraic Coding Theories.* Salvatore D. Morgera, Hari Krishna. Academic Pr, 1989, xx + 233 pp, $37.95. [ISBN: 0-12-506995-2] The intent of this book is to explore the design of computationally efficient digital signal processing algorithms over finite fields and the relation of these algorithms to algebraic error correcting codes. Assumes familiarity with abstract algebra and algebraic coding theory. Includes references but not exercises. CEC

Applications (Economics), S(18). *Econometrics of Structural Change.* Ed: Walter Krämer. Stud. in Empirical Econ. Physica-Verlag (US Distr: Springer-Verlag), 1989, ix + 128 pp, $50.60. [ISBN: 0-387-91357-2] A collection of nine papers on various aspects of structural change (coefficients changing over time) in econometric time series models. The Chow test to look for a change in coefficients at a single point in time is analyzed. Random-walk models, Bayesian analysis, sequential approaches, transformations, goodness-of-fit tests, and CUSUM tests are looked at in various papers. A 20-page annotated bibliography containing about 400 entries concludes the book. JJ

Applications (Engineering), P. *Machine Vision for Inspection and Measurement.* Ed: Herbert Freeman. Perspect. in Comput., V. 24. Academic Pr, 1989, x + 320 pp, $39.95. [ISBN: 0-12-266719-0] Eleven papers on applications ranging from integrated circuit inspection to identifying defects in glass to making precise measurements of the curvature of the human cornea. BC

Applications (Engineering), P. *Lecture Notes in Control and Information Sciences-132: Interactive Multi-Objective Programming as a Framework for Computer-Aided Control System Design.* W.-Y. Ng. Springer-Verlag, 1989, xv + 182 pp, $24.90 (P). [ISBN: 0-38-51504-6] This research monograph provides some updated results of research into the areas of Computer-Aided Design (CAD). It then reports on a new CAD technique called Interactive Multi-Objective Programming (IMOP) which enhances the cooperation and interaction between a human designer and the computer software. IMOP lets the designer have more control and more flexibility in providing the parameters to the design process. GMS

Applications (Fluid Dynamics), P. *Lecture Notes in Mathematics-1385: Relativistic Fluid Dynamics.* Eds: A. Anile, Y. Choquet-Bruhat. Springer-Verlag, 1989, v + 308 pp, $28.60 (P). [ISBN: 0-387-51466-X] Lectures from the first 1987 session of the Centro Internazionale Matematico Estivo. Four main lectures on covariant theory of conductivity, Hamiltonian methods, covariant fluid mechanics and thermodynamics, and relativistic plasmas are followed by five shorter papers. JO

Applications (Physical Science), P. *Transport Theory, Invariant Imbedding, and Integral Equations.* Eds: Paul Nelson, *et al.* Lect. Notes in Pure & Appl. Math., V. 115. Marcel Dekker, 1989, xxvi + 444 pp, $110 (P). [ISBN: 0-8247-8158-9] Twenty-seven papers from a conference honoring G. Milton Wing. The technical contributions are primarily for specialists, but Wing's short paper, "Counting Backward from 10," should be read by everyone. BC

Applications (Physical Science), T(16-18: 1), S. *Mathematical Problems from Combustion Theory.* Jerrold Bebernes, David Eberly. Appl. Math. Sci., V. 83. Springer-Verlag, 1989, 177 pp, $34. [ISBN: 0-387-97104-1] The analysis of the underlying mathematical models is not only purposeful and thorough, but reflects an esthetic appreciation not usually found in texts on applied mathematics. MU

Applications (Physics), P. *Group Theory in Physics, Volume III: Supersymmetries and Infinite-Dimensional Algebras.* J.F. Cornwell. Techniq. of Physics, V. 10. Academic Pr, 1989, xxii + 628 pp, $55. [ISBN: 0-12-189805-9] Super-objects are, roughly, objects with a notion of grading added. This volume presents the theories of Lie and Poincaré superalgebras and supergroups, and of Virasoro and Kac-Moody superalgebras, as well as algebraic as-

pects of strings and superstrings. JO

Applications (Physics), T(18), P. *Aspects of Quantum Field Theory in Curved Space-Time.* Stephen A. Fulling. Math. Soc. Stud. Texts, V. 17. Cambridge University Pr, 1989, ix + 315 pp, $49.50; $19.95 (P). [ISBN: 0-521-34400-X; 0-521-37768-4] Discusses wave equations and quantum fields in curved space-times. Begins with background in general quantum theory and in eigenfunction expansions for second-order elliptic differential operators. MLR

Applications (Physics), T(18: 1), S, P. *Finite Quantum Electrodynamics.* G. Scharf. Texts & Mono. in Physics. Springer-Verlag, 1989, x + 224 pp, $49.50. [ISBN: 0-387-51058-3] Written to avoid the pedagogical flaws in the standard approach to this topic, this high-level text will unfortunately be virtually inaccessible to all but the expert. MU

Applications (Physics), P. *Lecture Notes in Control and Information Sciences-133: Stochastic Models for Laser Propagation in Atmospheric Turbulence.* R.P. Leland. Springer-Verlag, 1989, vii + 145 pp, $24.90 (P). [ISBN: 0-387-51538-0] Mathematically rigorous analysis of function space models for laser propagation in turbulence. Simulation results support theory. TH

Applications (Physics), S(18), P. *Lecture Notes in Mathematics-1396: Quantum Probability and Applications IV.* Eds: L. Accardi, W. von Waldenfels. Springer-Verlag, 1989, vi + 355 pp, $33 (P). [ISBN: 0-387-51613-1] Proceedings of the Year of Quantum Probability held at the University of Rome II, 1987. MU

Applications (Physics), T(16-17: 1, 2), L. *An Introduction to Hilbert Space and Quantum Logic.* David W. Cohen. Problem Books in Math. Springer-Verlag, 1989, xii + 149 pp, $39.80. [ISBN: 0-387-96870-9] This unusual text assumes little of the reader other than the maturity attained by upper-level undergraduate mathematics students. Beginning with a quick tour of measure theory and Hilbert space, the book culminates in leading the reader into the Einstein-Podolsky-Rosen Paradox. A very worthwhile intellectual excursion. MU

Applications (Physics), P. *Noise in Nonlinear Dynamical Systems, Volume 3: Experiments and Simulations.* Eds: Frank Moss, P.V.E. McClintock. Cambridge University Pr, 1989, xvi + 278 pp, $75. [ISBN: 0-521-35265-7] Third of three volumes comprising a collection of research papers by different authors which may be read individually. This last volume contains nine articles dealing with the experimental aspects of the field. KS

Applications (Physics), T(17-18: 1, 2), S, P, L. *Gauge Field Theories.* Stefan Pokorski. Mono. on Math. Physics. Cambridge U Pr, 1989, xiii + 394 pp, $34.50 (P). [ISBN: 0-521-36846-4] An advanced textbook (with exercises) covering path integrals, Feynman rules, renormalization, QED and QCD, chiral symmetry and anomalies, symmetry breaking, and supersymmetry (1987 hardcover edition, TR, January 1988). BC

Applications (Physics), P. *Conformal Invariance and String Theory.* Ed: Petre Dita, Vladimir Georgescu. Perspect. in Physics. Academic Pr, 1989, xiii + 557 pp, $59.95. [ISBN: 0-12-218100-X] Sixteen papers on conformal invariance in statistical mechanics and its relationships with string theory. Yu. I. Manin summarizes the situation well on page 293: "The development of theoretical physics in the last quarter of the twentieth century is guided by a very romantic system of values." BC

Applications (Physics), T(17-18: 1), S, P. *Principles of Statistical Radiophysics 3: Elements of Random Fields.* S.M. Rytov, Yu. A. Kravtsov, V.I. Tatarskii. Springer-Verlag, 1989, x + 239 pp, $99. [ISBN: 0-387-17829-5] Basic definitions, properties, and forms of random fields. Intended as a text for radio physicists and engineers. Includes solved exercises. Note the price. BC

Applications (Physics), T(18: 2),S, P. *Analysis, Manifolds and Physics, Part II: 92 Applications.* Yvonne Choquet-Bruhat, Cécile DeWitt-Morette. North-Holland (US Distr: Elsevier Science, 1989, xii + 449 pp, $40. [ISBN: 0-444-87071-7] Organized under the following topics: review of fundamental notions of analysis; differential calculus on Banach spaces; differentiable manifolds; integration on manifolds; Riemannian manifolds; Kählerian manifolds; connections on a principal fibre bundle; and distributions (1977 edition, TR, May 1979). MU

Applications (Physics), T*(15), S, L. *Space-time and Singularities, An Introduction.* Gregory L. Naber. London Math. Soc. Stud. Texts, V. 11. Cambridge U Pr, 1988, ix + 178 pp, $39.50; $16.95 (P). [ISBN: 0-521-33327-X; 0-521-33612-0] Hawking's Theorem (paraphrased): Any universe which is stably causal, deterministic, expanding, and has attracting gravity, must be singular. Here, the existence of a singularity is indicated by being geodesicly incomplete. This book is single-minded: to lead the reader to understand and (maybe) prove Hawking's Theorem, all else is sacrificed. The requirements are first courses in real analysis, linear algebra, and physics. This could be a fun book for some motivated mathematics majors. MPR

Reviewers

DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; WE: William Etter, Macalester; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; RWJ: Roger W. Johnson, Carleton; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; JO: Jeff Ondich, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, St. Olaf; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; JAS: J. Arthur Seebach, Jr., St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

THE MATHEMATICAL SCIENCES

Games and Economic Behavior

Editor

Ehud Kalai

Northwestern University, Evanston, Illinois

Games and Economic Behavior publishes original and survey papers dealing with game-theoretic modeling in the social, biological, and mathematical sciences. Papers published are mathematically rigorous as well as accessible to readers in related fields. The purpose of the journal is to facilitate cross-fertilization between the theory and application of game-theoretic reasoning.

Research Areas Include

- Game theory
- Biology
- Economics
- Computer science
- Political science
- Mathematics
- Psychology

A Representative Selection of Articles

Robert J. Aumann and Sylvain Sorin
Cooperation and Bounded Recall

Itzhak Gilboa and Eitan Zemel
Nash and Correlated Equilibria: Some Complexity Considerations

Nimrod Megiddo
On Computable Beliefs of Rational Machines

Alvin E. Roth
Two-Sided Matching with Incomplete Information about Others' Preferences

Volume 2 (1990), 4 issues
ISSN 0899-8256

In the U.S.A. and Canada: $100.00
All other countries: $122.00

Historia Mathematica

Editor

Eberhard Knobloch

Technische Universität Berlin, West Germany

Managing Editor

David E. Rowe

Pace University, Pleasantville, New York

Historia Mathematica is concerned with the history of all aspects of the mathematical sciences in all parts of the world and all historical periods. The journal publishes occasional biographies of mathematicians and historians, studies of organizations and institutions, essays on historiography, and articles on the interactions among all facets of mathematical activity and other aspects of culture and society.

Published under the Auspices of the International Commission on the History of Mathematics of the Division of the History of Science of the International Union of the History and Philosophy of Science

Volume 17 (1990), 4 issues
ISSN 0315-0860

In the U.S.A. and Canada: $78.00
All other countries: $97.00



Sample copies and privileged personal rates are available upon request. For more information, please write or call:

ACADEMIC PRESS, INC., Journal Promotion Department
1250 Sixth Avenue, San Diego, CA 92101, U.S.A.
(619) 699-6742

All prices are in U.S. dollars and are subject to change without notice.

S0107



Wadsworth & Brooks/Cole Advanced Books & Software

Textbooks

If you wish to consider one of the following books for adoption in a course you are teaching, please write for a complimentary review copy on department letterhead to the address below.

Computability: Computable Functions, Logic, and the Foundations of Mathematics
by Richard L. Epstein and Walter A. Carnielli
An undergraduate introduction to the mathematics of computable functions and logic using readings from Hilbert, Gödel, Turing, Post, Church, and many others.
1989. 297 pages. Casebound.
ISBN 0-534-10356-1.

APL with a Mathematical Accent
by Clifford A. Reiter and William R. Jones
Uses examples from the mathematical sciences to motivate the development of APL usage. Independent chapters allow the book to be used as the main text or a supplement.
1990. 225 pages. Paperbound.
ISBN 0-534-12864-5.

Statistics: Theory and Methods
by Donald A. Berry and Bernard W. Lindgren
Undergraduate mathematical statistics text designed for students with relatively weak calculus backgrounds. Many of the applications use real data. 1990. 690 pages. Casebound. ISBN 0-534-09942-4.

Complex Variables, Second Edition
by Stephen D. Fisher
New edition of a popular undergraduate text that presents both the theory and applications of complex variables.
April 1990. 424 pages. Casebound.
ISBN 0-534-13260-X.

Real Analysis and Probability
by R.M. Dudley
Graduate text that offers an exposition of modern probability theory and of the interplay between the properties of metric spaces and probability measures. 1989. 448 pages. Casebound. ISBN 0-534-10050-3.

Software and Professional Reference Books

Order our books and/or software using our toll-free number (800) 354-9706, or write to us at the address below. All our products are backed by a 30-day money-back guarantee.

Graphical Aids for Stochastic Processes
by Bob Fisch and David Griffeath
Winner of the 1988 EDUCOM-NCRIPTAL Award for Best Mathematics Software
Six-disk educational software package for the IBM PC that provides a colorful introduction to probability and random processes.
1988. For IBM PC/compatibles with color graphics card/320K RAM.
ISBN 0-534-09042-7. **$105.**

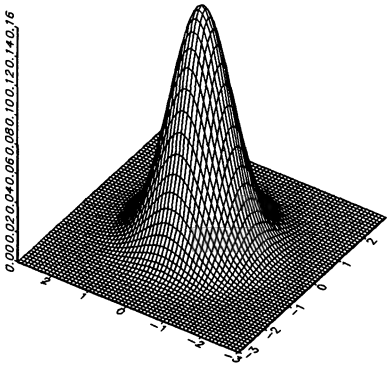
Statable: Electronic Tables for Statisticians and Engineers, Version 1.0
by Cytel Software Corporation
With just a few keystrokes, the tail area or percentage point you want for the twenty-five most commonly used statistical distributions appears in a pop-up window. Eliminates hunting for books of tables, interpolation, and the possibility of errors in calculation. Can run by itself or as a memory-resident program.
1989. One IBM PC disk and 32 pages documentation. Paperbound.
5 1/2" disks: ISBN 0-534-11964-6.
3 1/2" disks: ISBN 0-534-11965-4. **$39.95.**

A Dictionary of Real Numbers
by Jonathan M. Borwein and Peter B. Borwein
For researchers who have encountered various numbers computationally and want to know if these numbers have some simple form. This reference book lists over 100,000 eight digit real numbers in the interval [0,1) that arise as the first eight digits of special values of familiar functions. January 1990. 470 pp. Cloth. ISBN 0-534-12480-8. **$69.95.**

JMK009
Wadsworth & Brooks/Cole
Advanced Books & Software
Pacific Grove, CA 93950 (408) 373-0728

Can Your Word Processor Show You This Page on the Screen?

$$W^{3\beta}_{m_1 n_1 n_2}(p_1,p_2) = U^{3\beta}_{m_1 n_1}(p_1,p_2) + \int_0^\infty \frac{dp_3 p_3^2}{8\pi^3} \sum_n \sum_m \sum_{\alpha_2} \sum_{\beta_2} \sum_{n'} \sum_{n''} (-1)^m$$
$$\times \left(\frac{U^{33}_{m_1 n_1}(p_1,p_2)}{p_3^2 - k^2} \right) z_{3 m_1 n_1} h_n(p_3,p_2) \cdot \alpha^{\alpha_2 3}_{mn(m_1-m)n' n_2} \alpha^{\beta_2 \beta}_{-mn(-m_1+m)n'' n_2} W^{\alpha_2 \beta_2}_{(m_1-m)n'n''}(p_3,p_2).$$



EXP: The Scientific Word Processor, Version 2.0
Simon L. Smith
The page you see here was written with **EXP** 2.0, printed at 300 dpi, and reduced to size. This page shows the actual screen display you see while editing this document, not a special "page-preview" mode. The graph was drawn by a graphing program and then imported into EXP.

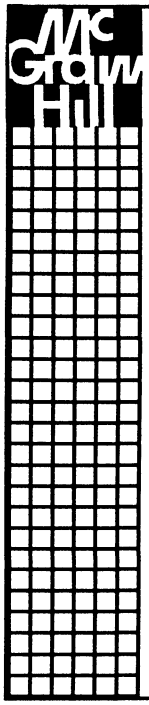
EXP remains the fastest and truest "what you see is what you get" scientific word processor for the IBM PC, featuring automatic sizing and centering of many symbols; macros; windows; and automatic numbering. Version 2.0 includes the ability to import bit-mapped graphics that can be seen on the screen as they will print; spelling checker; increased macro capabilities; more automatically scaling symbols; automatic conversion of variables into italics; ability to export files to popular desktop publishing programs; and support for VGA/HP DeskJet Plus.
February 1990. ISBN 0-534-11970-0. $295. Upgrade and site license prices available.

Mainframe Mathematical Power for the Macintosh
Maple, Version 4.2: Symbolic Computation for the Macintosh *by The Symbolic Computation Group, University of Waterloo*
Maple puts symbolic computation at your fingertips at a price you can afford! Put Maple's speed and reliability to work on Macintosh—SE, SE/30, Mac II series—even the Mac Plus. And **Maple** requires only one megabyte!

Maple combines symbolic computation and numerical computation with two-dimensional graphics to give you more power per dollar than any other symbolic processing software. **Maple's** unique library of more than 1500 built-in functions is fully extensible. The **Maple** package includes the Macintosh interface, complete **Maple** libraries with source code and help files, *First Leaves for the Macintosh* (a tutorial), and the *Maple Reference Manual.*
1989. ISBN 0-534-10218-2. $395.
Site license prices available.

A Powerful Software Tool for the IBM PC that Solves Differential Equations Fast
PhasePlane: The Dynamical Systems Tool, Version 3.0 *by Bard Ermentrout*
This interactive program for solving up to 20 ordinary differential equations, difference equations, and differential-delay equations allows users to • create self-running and interactive demos • choose from seven different numerical methods, including Gear's method for "stiff" equations • draw nullclines, direction fields, and global flows for two-dimensional models • create 3-D graphics, perspective transformations, and color coding of trajectories by velocity or amplitude. **PhasePlane** includes a built-in scientific calculator and plotter/LaserJet support.
January 1990. ISBN 0-534-12894-7. $42.00.
Site license prices available.

For U.S. orders, call (800)354-9706.
To request a copy of our Advanced Books & Software catalog, please write:
Wadsworth & Brooks/Cole
Advanced Books & Software
511 Forest Lodge Road
Pacific Grove, CA 93950



The Mathematical Symbol For Quality...

with a full complement of upper-level texts.

NEW FOR 1990!

COMPLEX VARIABLES WITH APPLICATIONS, 5/e
Churchill/Brown

EXERCISES IN MULTIVARIABLE & VECTOR CALCULUS
Curjel

INTRODUCTION TO PROBABILITY AND STATISTICS: PRINCIPLES AND APPLICATIONS FOR ENGINEERING AND THE COMPUTER SCIENCES, 2/e
Milton/Arnold

MULTIVARIATE STATISTICAL METHODS, 3/e
Morrison

A BOOK OF ABSTRACT ALGEBRA, 2/e
Pinter

CALCULUS: AN INTEGRATED APPROACH
Small/Hosack

For more information, contact your local sales representative or write: McGraw-Hill Publishing Company, Comp Processing and Control, P.O. Box 448, Hightstown, NJ 08520.



Fields& Operators

Surfaces. Vector Fields. Differential Operators. In two or three dimensions.

And it's animated!

From our mail...

"Another Success from Lascaux Graphics!"
M.E. City College of N Y

"Excellent Program!! ..An exceptional educational aid." D.W. - Dow Chemical

"Just what I've been looking for!"
D.B. Olivet College

"Keep up the valuable programs at decent prices."
D.W.

"Amazing Software - write more." Dr. R. B. TNU
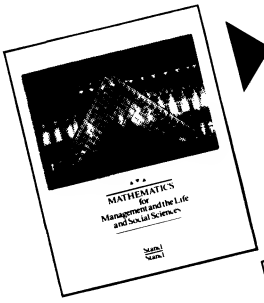
"Without your program's powerful layering and animation capabilities, we would not have attempted such a task." M. T. Univ. of Wisconsin

"A super product "
Recreational & Educational Computing

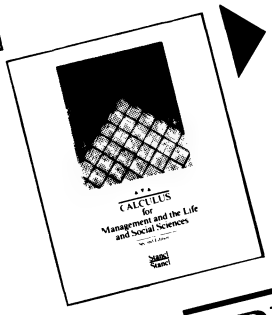
For PC or Macintosh

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

IRWIN-
The NEW FORCE
in Mathematics...



MATHEMATICS
for
Management and the Life and Social Sciences



CALCULUS
for
Management and the Life and Social Sciences
Second Edition

All Available Now!



BRIEF CALCULUS
for
Management and the Life and Social Sciences
Second Edition

Stancl & Stancl

To request examination copies of these texts, contact your IRWIN representative or call 1-800-323-4560 (In Canada, 1-800-268-4178, in Ontario, 1-416-298-1588.)

IRWIN

1818 Ridge Road Homewood, IL 60430



Random Walks and Electric Networks,

by J. Laurie Snell and Peter Doyle

xiii + 159 pages. Hardbound
List: $25.00 MAA Member: $19.00

In this newest addition to the Carus Mathematical Monographs, the authors examine the relationship between elementary electric network theory and random walks, at a level which can be appreciated by the able college student. We are indebted to them for presenting this interplay between probability theory and physics in so readable and concise a fashion.

Central to the book is Polya's beautiful theorem that a random walker on an infinite street network in d-dimensional space is bound to return to the starting point when d = 2, but has a positive probability of escaping to infinity without returning to the starting point when d = 3. The authors interpret this theorem as a statement about electric networks, and then prove the theorem using techniques from classical electrical theory. The techniques referred to go back to Lord Rayleigh who introduced them in connection with an investigation of musical instruments.

In Part I the authors restrict themselves to the study of random walks on finite networks, establishing the connection between the electrical concepts of current and voltage and corresponding descriptive quantities of random walks regarded as finite state Markov chains. Part II deals with the idea of random walks on infinite networks.

Table of Contents

***Part I:* Random Walks on Finite Networks**
Random Walks in One Dimension
Random Walks in Two Dimensions
Random Walks on More General Networks
Rayleigh's Monotonicity Law

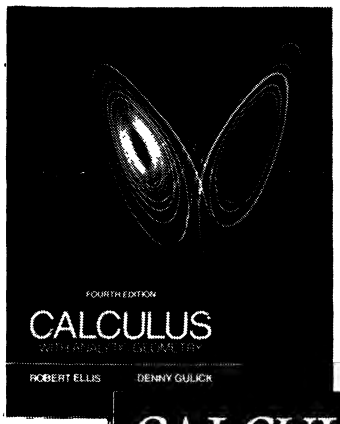
***Part II:* Random Walks on Infinite Networks**
Pólya's Recurrence Problem
Rayleigh's Short-Cut Method
The Classical Proofs of Polya's Theorem
Random Walks on More General Infinite Networks

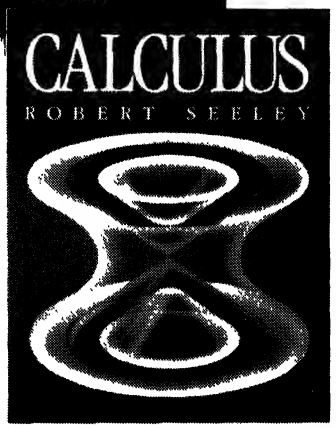


Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

FOR YOUR 1990 TEXTBOOK NEEDS.

CALCULUS WITH ANALYTIC GEOMETRY
Fourth Edition
Robert Ellis
Denny Gulick
Hardcover, 1990.
Just Published!





CALCULUS
Robert Seeley
Hardcover, 1990.
Just Published!

PRECALCULUS
Functions and Graphs
Second Edition
Bernard Kolman
Arnold Shapiro
Hardcover, 1990.
Just Published!

LINEAR ALGEBRA
Third Edition
Michael O'Nan
Herbert B. Enderton
Hardcover, 1990.
Just Published!

INTRODUCTORY COMBINATORICS
Second Edition
Kenneth P. Bogart
Hardcover, 1990.
Just Published!

HARCOURT BRACE JOVANOVICH, INC.
College Sales Office
7555 Caldwell Avenue,
Chicago, IL 60648
(312) 647-8822

HBJ

EXPLORING SMALL GROUPS
A Tool for Learning Abstract Algebra
(software)
Ladnor D. Geissinger
Disk and Manual, 1989.
Just Published!

MAA STUDIES IN MATHEMATICS

Studies in Combinatorics

Gian-Carlo Rota, editor
Volume #17, MAA Studies in Mathematics
272 pp., 1978, Hardbound, ISBN-0-88385-177-2
List: $27.50 MAA Member: $21.50

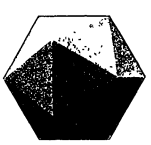
"This is a gem of a book. In layout and size it brings back memories of the classic **Combinatorial Mathematics** *by H.J. Ryser. The index is very good and the editor is to be warmly congratulated on bringing out a book of enormous value and uniformly high standard. It is the sort of book which could be used with great profit as the basis of a weekly working seminar on combinatorics. I recommend it unreservedly."*

D.J.A. Welsh in ***Bulletin of the Institute of Mathematics and its Applications***

This excellent book in the **MAA Studies in Mathematics** series contains seven papers by prominent researchers in combinatorics. Although written by different authors, the papers have the unity of chapters in a monograph. Together they present a fairly comprehensive exposition of some recent developments in combinatorics. Much of the material has application outside mathematics to computer science, operations research and network and circuit theory. Begun at the time of Euler, combinatorics has become one of the most active fields in mathematics. A significant feature of **Studies in Combinatorics** is its accessibility to any mathematician qualified to teach at the college level, as well as to a great many students. This volume is also accessible to a wide scientific public outside mathematics.

Table of Contents

Combinatorial Matrix Theory H.J. Ryser
Proof Techniques in the Theory of Finite Sets Curtis Greene and Daniel J. Kleitman
Ramsey Theory R.L. Graham and B.L. Rothschild
Generating Functions Richard P. Stanley
Nonconstructive Methods in Discrete Mathematics Joel Spencer
Matroids and Combinatorial Geometries Tom Brylawski and Douglas G. Kelly
Combinatorial Constructions Marshall Hall, Jr.



Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Prindle, Weber and Schmidt

Celebrating our Silver Anniversary in 1990!



The Prindle, Weber and Schmidt Series in Mathematics is both pleased and honored to announce the celebration of its Silver Anniversary. Over the past 25 years we've dedicated ourselves to leading the industry in innovative texts and responsive service. Our commitment to the discipline and its changing needs is as strong as ever and even today we are looking toward the future with our eyes and ears wide open to new ideas and trends. We look forward to servicing you, the mathematics educators of America and the world, for another 25 years and upholding our distinguished reputation as a proven leader in mathematics textbook publishing.

New titles for 1990!

DEVELOPMENTAL MATHEMATICS, 4/E
Wood, Capell and Hall

GEOMETRY AND LOGIC CONCEPTS: A Supplement for Developmental Mathematics
Kaufmann

COLLEGE ALGEBRA, 2/E
Kaufmann

COLLEGE ALGEBRA AND TRIGONOMETRY, 2/E
Kaufmann

TRIGONOMETRY
Gantner and Gantner

PRECALCULUS: FUNCTIONS AND GRAPHS, 6/E
Swokowski

APPLIED FINITE MATHEMATICS, 3/E
Tan

CALCULUS FOR THE MANAGERIAL, LIFE, AND SOCIAL SCIENCES, 2/E
Tan

APPLIED CALCULUS, 2/E
Tan

FOUNDATIONS AND FUNDAMENTAL CONCEPTS OF MATHEMATICS, 3/E
Eves

ELEMENTARY LINEAR ALGEBRA WITH APPLICATIONS, 2/E
Nicholson

INTRODUCTION TO REAL ANALYSIS
Brabenec

APPLIED PARTIAL DIFFERENTIAL EQUATIONS
Trim

To receive a complimentary review copy of these or any other outstanding titles from the Prindle, Weber and Schmidt Series in Mathematics, call:

TOLL-FREE:

1-800-343-2204

(in MA 617-542-3377)



PWS-KENT Publishing Company
20 Park Plaza
Boston, MA 02116
A Division of Wadsworth, Inc.
Partners in Education

CAMBRIDGE UNIVERSITY PRESS

LECTURES ON MINIMAL SURFACES
Volume I
Johannes C.C. Nitsche
1989/563 pp./24427-7/Hardcover $135.00

INTRODUCTION TO NON-COMMUTATIVE NOETHERIAN RINGS
K. Goodearl and R. Warfield Jr.
London Mathematical Society Student Texts 16
1989/300 pp./36086-2/Hardcover $49.50
36925-8/Paper $19.95

ALGORITHMIC ALGEBRAIC NUMBER THEORY
M. Pohst and H. Zassenhaus
Encyclopedia of Mathematics and its Applications 30
1989/480 pp./33060-2/Hardcover $89.50

ASPECTS OF QUANTUM FIELD THEORY IN CURVED SPACETIME
Stephen A. Fulling
London Mathematical Society Student Texts 17
1989/300 pp./34400-X/Hardcover $49.40
37768-4/Paper $19.95

ALGEBRAIC HOMOTOPY
Hans Joachim Baues
Cambridge Studies in Advanced Mathematics 15
1989/480 pp./33376-8/Hardcover $89.50

MATHEMATICAL MODELLING
A Case Study Approach
R.R. Clements
1989/166 pp./34340-2/Hardcover $44.50

NUMBER THEORY AND DYNAMICAL SYSTEMS
Edited by M. Dodson and J. Vickers
London Mathematical Society Lecture Note Series
1989/172 pp./36919-3/Paper $24.95

ADVANCES IN HOMOTOPY THEORY
Papers in Honour of I.M. James, Cortona 1988
Edited by S. Salamon, B. Steer and W. Sutherland
London Mathematical Society Lecture Note Series 139
1989/177 pp./37907-5/Paper $27.95

ELEMENTS OF FUNCTIONAL ANALYSIS
I.J. Maddox
1989/200 pp./35350-5/Hardcover $59.50
35868-X/Paper $22.95

THE THEORY AND APPLICATIONS OF HARMONIC INTEGRALS
W.V.D. Hodge
Foreword by Sir Michael Atiyah
Cambridge Mathematical Library
1989/281 pp./35881-7/Paper $22.95

Now in paperback...

ERGODIC THEORY
Karl Petersen
Cambridge Studies in Advanced Mathematics
1983/352 pp. /38997-6/Paper $29.95

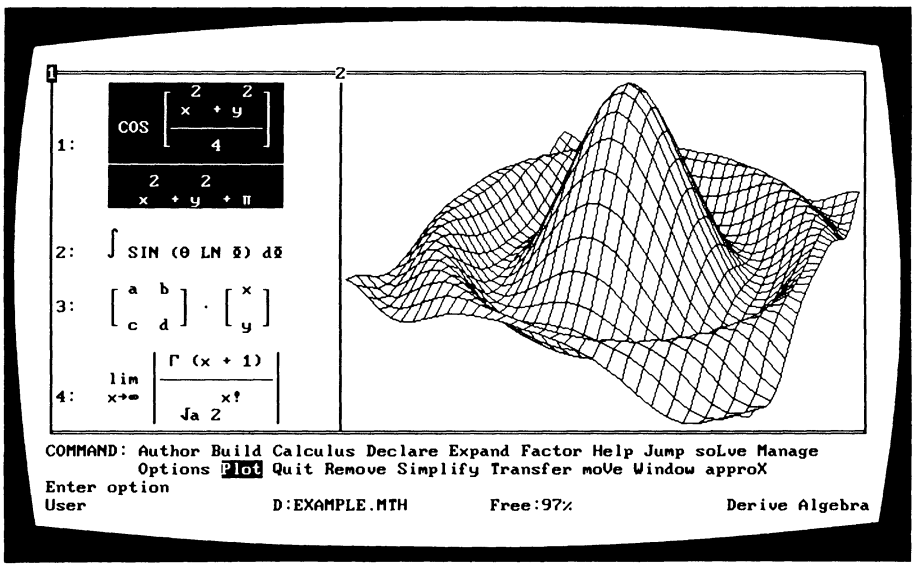
REGULAR VARIATION
N.H. Bingham, C.M. Goldie and J.L. Teugels
Encyclopedia of Mathematics and Its Applications 27
1987/512 pp./37943-1/Paper $34.50

At bookstores or order form

CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, NY, NY 10011. Call toll-free: 800-872-7423, outside NY State. 800-227-0247, NY State only.
MasterCard & VISA accepted. Prices subject to change.

DERIVE

A Mathematical Assistant



2000 years of mathematical knowledge on a disk

DERIVE, the successor to **muMATH**, is a powerful computer algebra system for your PC compatible computer that provides the following capabilities:

- Exact and approximate arithmetic to thousands of digits
- Equations, complex numbers, trigonometry, calculus, vectors, and matrices
- 2D and 3D function plotting with zooming capability
- MDA, CGA, EGA, VGA, and Hercules graphics and text support
- Attractive 2D mathematical display of formulas
- Easy to use menu-driven interface with on-line help
- Ideal for engineers, scientists, students and teachers
- $200 plus shipping: **Call or write** for information.

System requirements: IBM PC or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (360K) or a 3½ inch (760K) diskette drive. **Or** NEC PC-9801 or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (640K) diskette drive.

DERIVE and muMATH are trademarks of Soft Warehouse, Inc. Hercules is a trademark of Hercules Computer Technology, Inc. IBM is a registered trademark of International Business Machines Corp. MS-DOS is a registered trademark of Microsoft Corp. NEC is a registered trademark of Nippon Electric Company.

Soft Warehouse INC
3615 Harding Avenue, Suite 505 • Honolulu, HI 96816
(808) 734-5801 after noon PST
Handcrafted software for the mind.

©1988 Soft Warehouse, Inc.

QUALITY TEXTBOOKS

Spinors and Calibrations

F. Reese Harvey
Rice University, Houston, Texas

Progress in mathematics is based on a thorough understanding of the mathematical objects under consideration, and yet many books proceed to discuss general statements and assume that the reader can and will provide the mathematical infrastructure of examples and counterexamples. This book makes a deliberate effort to correct this situation: it is a collection of examples. The following table of contents describes its breadth and reveals the underlying motivation—differential geometry—in its many facets: Riemannian, symplectic, Kähler, hyper-Kähler, as well as complex and quaternionic.

CONTENTS: Classical Groups I. The Eight Types of Inner Product Spaces. Classical Groups II. Euclidean/Lorentzian Vector Spaces. Differential Geometry. Normed Algebras. Calibrations. Matrix Algebras. **Spinors:** The Clifford Algebras. The Groups Spin and Pin. The Clifford Algebras C (r,s) as Algebras. The Split Case C (p,p). Inner Products on the Spaces of Spinors and Pinors. Low Dimensions.

December 1989, c. 336 pages, $39.95 (tentative)
ISBN: 0-12-329650-1

The Numerical Solution of Ordinary and Partial Differential Equations

Granville Sewell
University of Texas at El Paso

"This book is excellent....The presentation is clear and well motivated....Sewell never includes theory for its own sake; instead, he constantly draws the reader's attention to its practical value."

—MATHEMATICS OF COMPUTING

Introducing the use of finite difference and finite element methods for solving ordinary and partial differential equations, this book can be used as a textbook for graduate or advanced undergraduate courses.

1988, 271 pages, $39.95
ISBN: 0-12-637475-9

Finite Element Analysis for Undergraduates

J. Ed Akin
Rice University, Houston, Texas

This text is intended to introduce the topic of finite element analysis to undergraduates. It requires the minimum amount of prerequisite material. These include the first courses in calculus, physics, and mechanics of solids. An introduction to matrix algebra is included. Some FORTRAN programs have been included to illustrate the implementation of matrix algebra.

In Paperback: $29.50/ISBN: 0-12-047656-8
Casebound: $58.00/ISBN: 0-12-047655-X
1987, 336 pages

Numerical Methods for Initial Value Problems in Ordinary Differential Equations

Simeon Ola Fatunla
University of Benin, Nigeria

The first half of this book deals with Taylor, Runge Kutta, linear one-step, linear multistep, multiderivative, extrapolation, and predictor corrector methods, which makes it suitable for one-semester courses for seniors and graduates.

1988, 295 pages, $44.50
ISBN: 0-12-249930-1



ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
Book Marketing Department #S03020
1250 Sixth Avenue, San Diego, CA 92101

CALL TOLL FREE
1-800-321-5068
Quote this reference number for *free postage and handling* on your prepaid order ➥ **S03020**
or
Call 619-699-6400 for textbook adoption information.

Prices subject to change without notice. ©1989 by Academic Press, Inc
All Rights Reserved TS —S03020

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 3 March 1990

Contents

(ISSN 0002-9890)

ARTICLES

An Alternative Proof of the Lindemann-Weierstrass Theorem
FRITS BEUKERS, JEAN-PAUL BÉZIVIN, AND PHILIPPE ROBBA 193

Lucas's Theorem and Some Related Results for Extended Pascal Triangles RICHARD BOLLINGER AND CHARLES BURCHARD 198

The Remainder in Taylor's Formula ESTEBAN I. POFFALD 205

LETTERS TO THE EDITOR 214

NOTES

A Link Between the Jordan Curve Theorem and the Kuratowski Planarity Criterion CARSTEN THOMASSEN 216

A New Series Representation for $\zeta(3)$ JOHN A. EWELL 219

The Dirichlet Problem for a Disk DAVID MINDA 220

Diagonalization over Commutative Rings R. BRUCE RICHTER AND WILLIAM P. WARDLAW 223

THE TEACHING OF MATHEMATICS

Random Triangles JOE WHITTAKER 228

A New Extension of the Derivative DANIEL KOPEL AND MICHAEL SCHRAMM 230

Remainder Estimates in Taylor's Theorem G. B. FOLLAND 233

Symmetry in Logos and Hubcaps JOSEPH A. GALLIAN 235

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 239

Advanced Problems and Solutions 253

REVIEWS

A Formalization of Set Theory without Variables by Alfred Tarski and Steven Givant LISL GAAL 258

Complex Analysis by Joseph Bak and Donald J. Newman. Invitation to Complex Analysis by Ralph Philip Boas LAWRENCE ZALCMAN 262

Fractals Everywhere by Michael Barnsley ANTHONY BARCELLOS 266

TELEGRAPHIC REVIEWS 269

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

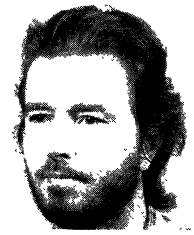
PRINTED IN THE UNITED STATES OF AMERICA

An Alternative Proof of the Lindemann-Weierstrass Theorem

F. BEUKERS, J. P. BÉZIVIN, P. ROBBA

Dedicated to the Memory of Philippe Robba

FRITS BEUKERS: born in 1953. University education in Leiden, Netherlands. Currently teaching at the State University of Utrecht. Research interests are number theory and arithmetic properties of linear differential equations.



JEAN-PAUL BÉZIVIN: educated at the École Normale Supérieure and the Universities of Paris VI and VII. He currently teaches at the University of Paris VI. Research interests are p-adic analysis, linear recurrent sequences and arithmetic properties of formal power series.



PHILIPPE ROBBA: born in 1941, died on October 12, 1988. Doctorate degree at the University of Paris in 1972, and professor at the University Paris XI from 1975 until his untimely death. Domain of research: p-adic analysis, mainly p-adic differential equations with emphasis on p-adic cohomologies.



Introduction. In December 1987 J. P. Bézivin and Ph. Robba found a new proof of the Lindemann-Weierstrass theorem as a by-product of their criterion of rationality for solutions of differential equations. Let us recall the Lindemann-Weierstrass theorem, to which we shall refer as LW from now on.

Let $\alpha_1, \ldots, \alpha_t, b_1, \ldots, b_t$ be algebraic numbers such that the b_i are all nonzero and the α_i are mutually distinct. Then

$$b_1 e^{\alpha_1} + b_2 e^{\alpha_2} + \cdots + b_t e^{\alpha_t} \neq 0.$$

It is well known that the transcendence of π follows from LW in the following way. Suppose, on the contrary, that π is algebraic. Then so is $\pi\sqrt{-1}$ and LW now implies that $e^{\pi\sqrt{-1}} + 1 \neq 0$, which is certainly not true. Thus we conclude that π is transcendental.

The usual proof of LW is essentially due to Hilbert and has been polished by a number of authors. One such version can be found in [**2**, Ch. XI] or in [**4**, Ch. I], [**3**]. The new proof of Bézivin and Robba looks totally different. It can be considered as

a direct consequence of their criterion of rationality for solutions of linear differential equations.

The Bézivin-Robba criterion is based on a theorem of Pólya-Bertrandias, which is far from easy and relies heavily on p- adic analysis. We refer the interested reader to [1]. In March 1988 F. Beukers found that the use of the Pólya-Bertrandias criterion is much too heavy and can be avoided in a very elementary way. The result is a new proof of LW which is elementary and can compete with the usual one in shortness and simplicity. There is always a possibility that the similarity between the proofs is stronger than one would expect at first sight. In fact, very soon after a first draft of this paper was written (May 1988) Yu. Nesterenko pointed out to us that the numbers $v_n(k)$ which we use are equal to the integrals

$$\sum_{j=1}^{t} b_j e^{\alpha_j} \int_{\alpha_j}^{\infty} e^{-x} x^{n-kt} (x-\alpha_1)^k \cdots (x-\alpha_t)^k \, dx$$

which are used in the Hilbert proof. In spite of such similarities we feel that the arguments of our proof are nice enough to present in front of a wider audience. We would like to thank the referee for several improvements upon our presentation.

THEOREM. *Let* $b_1, \ldots, b_t$, $\alpha_1, \ldots, \alpha_t \in \overline{\mathbb{Q}}$ *such that* $b_i \neq 0$ $\forall i$ *and the* α_i *are mutually distinct. Then*

$$b_1 e^{\alpha_1} + \cdots + b_t e^{\alpha_t} \neq 0.$$

Proof. Consider the Taylor series expansion

$$b_1 e^{\alpha_1 x} + \cdots + b_t e^{\alpha_t x} = \sum_{n=0}^{\infty} \frac{u_n}{n!} x^n,$$

where, clearly,

$$u_n = \sum_{i=1}^{t} b_i \alpha_i^n. \tag{1}$$

Put $(X-\alpha_1)\cdots(X-\alpha_t) = X^t - a_1 X^{t-1} - \cdots - a_t$. Clearly, for any $i = 1, \ldots, t$ and any $n \in \mathbb{Z}_{\geqslant 0}$

$$\alpha_i^{t+n} = a_1 \alpha_i^{t+n-1} + \cdots + a_t \alpha_i^n.$$

By taking suitable linear combinations and using (1) it follows that

$$u_{n+t} = a_1 u_{n+t-1} + \cdots + a_t u_n. \tag{2}$$

Without loss of generality we may assume that $u_n \in \mathbb{Q}$, $\forall n$. If not, then consider the product

$$\prod_{\sigma} \left(\sigma(b_1) e^{\sigma(\alpha_1)x} + \cdots + \sigma(b_n) e^{\sigma(\alpha_t)x} \right)$$

taken over all $\sigma \in \mathrm{Gal}(\mathbb{Q}(b_1, \ldots, b_t, \alpha_1, \ldots, \alpha_t)/\mathbb{Q})$, which, after evaluation, again acquires the form

$$\sum_i b_i' e^{\alpha_i' x},$$

where now the sets $\{b_i'\}$, $\{\alpha_i'\}$ are Galois-stable. This implies that the corresponding numbers a_i' and u_n' are rational. So from now on we assume $u_n \in \mathbb{Q}$, $\forall n$ and

$a_i \in \mathbb{Q}$ $(i = 1, \ldots, t)$. Let D be a common denominator of the a_i. Put $A = \max(1, |\alpha_i|)$. After multiplication with a suitable integer, if necessary, we may assume $u_0, \ldots, u_{t-1} \in \mathbb{Z}$. Hence, using (2) recursively,

$$D^n u_n \in \mathbb{Z}, \tag{3}$$

and by (1),

$$|u_n| \leqslant c_1 A^n$$

for some $c_1 > 0$ and all $n \geqslant 0$. Now suppose that $b_1 e^{\alpha_1} + \cdots + b_t e^{\alpha_t} = 0$, or, equivalently,

$$\sum_{r=0}^{\infty} \frac{u_r}{r!} = 0.$$

Put

$$v_n = n! \sum_{r=0}^{n} \frac{u_r}{r!}$$

and notice

$$|v_n| = n! \left| \sum_{r=0}^{n} \frac{u_r}{r!} \right| = n! \left| \sum_{r=n+1}^{\infty} \frac{u_r}{r!} \right| \leqslant \frac{c_1}{n+1} \sum_{r=n+1}^{\infty} \frac{A^r}{(r=n+1)!} = c_2 \frac{a^{n+1}}{n+1}. \tag{4}$$

If we had $A = D = 1$, like in the high school proof of $e \notin \mathbb{Q}$, inequality (4) gives us a contradiction since we have both $v_n \in \mathbb{Z}$ and $|v_n| \leqslant c_2/(n+1)$, i.e. $v_n = 0$ for all sufficiently large n, in other words $\sum_0^\infty v_n X^n$ is a polynomial. In our general case a similar principle works.

Claim:

$$\sum_{n=0}^{\infty} v_n X^n \in \mathbb{Q}(X).$$

Assuming the claim we proceed with the proof. Define

$$v(X) = \sum_{n=0}^{\infty} v_n X^n.$$

Notice that

$$\frac{v_n}{n!} - \frac{v_{n-1}}{(n-1)!} = \frac{u_n}{n!} \quad \text{or} \quad v_n - n v_{n-1} = u_n.$$

So,

$$\sum_{n=0}^{\infty} (v_n - n v_{n-1}) X^n = \sum_{n=0}^{\infty} u_n X^n. \tag{5}$$

Using (1), the right-hand side of (5) is seen to be

$$\sum_{n=0}^{\infty} u_n X^n = \sum_{i=1}^{t} \frac{b_i}{1 - \alpha_i X},$$

whereas the left-hand side equals

$$v(X) - X \frac{d}{dX}(X v(X)) = (1 - X) v(X) - X^2 \frac{d}{dX} v(X).$$

So (5) becomes

$$\mathscr{L}v(X) = \sum_{i=1}^{t} \frac{b_i}{1-\alpha_i X}, \qquad \mathscr{L} = -X^2\frac{d}{dX} + (1-X). \tag{6}$$

By the claim we know that $v(X) \in \mathbb{Q}(X)$ and so the non-zero poles of $\mathscr{L}v(X)$ have order at least two. However, the right-hand side of (6) has only simple poles. This contradiction proves our theorem, since the assumption $b_1 e^{\alpha_1} + \cdots + b_t e^{\alpha_t} = 0$ has turned out to be untenable. □

It now remains to prove our claim. We first observe that, as v is a solution of the differential equation (6), if v is a rational function then its poles must be at the points $1/\alpha_i$. Therefore we expect that there exists an integer k such that $(1 - a_1X - \cdots - a_tX^t)^k v(X)$ is a polynomial.

DEFINITION. *For any k, $n \in \mathbb{Z}_{\geqslant 0}$ we define $v_n(k)$ as coefficient in the formal power series*

$$\sum_{n=0}^{\infty} v_n(k)X^n = \left(1 - a_1X - \cdots - a_tX^t\right)^k \sum_{n=0}^{\infty} v_nX^n.$$

For later use we also note that

$$v_n(k+1) = v_n(k) - a_1v_{n-1}(k) - \cdots - a_tv_{n-t}(k) \quad \text{for all } n \geqslant t,\ k \geqslant 0. \tag{7}$$

LEMMA 1. *Let $C = 1 + |a_1| + \cdots + |a_t|$. For all $n \geqslant kt$ we have*

i) $|v_n(k)| \leqslant c_2A^nC^k$

ii) $D^nv_n(k) \in \mathbb{Z}$

iii) $k!$ *divides* $D^nv_n(k)$.

Proof. The first two assertions follow easily by induction on k from (3), (4) and (7). The third assertion can be shown as follows. Write

$$v_n = u_n + nu_{n-1} + \cdots + n(n-1)\cdots(n-k+2)u_{n-k+1} + w_n,$$

where $w_n = n!\sum_{r=0}^{\infty} u_r/r!$. Notice that $D^{n-k}w_n \in \mathbb{Z}$ and $k!$ divides $D^{n-k}w_n$. Consider the power series

$$v(X) = \sum_{n=0}^{\infty} v_nX^n \qquad w(X) = \sum_{n=0}^{\infty} w_nX^n.$$

Then

$$v(X) - w(X) = \sum_{n=0}^{\infty} \left\{u_n + nu_{n-1} + \cdots + n(n-1)\cdots(n-k+2)u_{n-k+1}\right\}X^n.$$

For any $0 \leqslant r \leqslant k-1$ we observe that

$$\begin{aligned}\sum_{n=0}^{\infty} n(n-1)\cdots(n-r+1)u_{n-r}X^n &= \sum_{i=1}^{t}\sum_{n=0}^{\infty} r!\binom{n}{r}b_i\alpha_i^nX^n \\ &= \sum_{i=1}^{t} r!\,b_i\alpha_i^rX^r\frac{1}{(1-\alpha_iX)^{r+1}} \\ &= \frac{P_r(X)}{\left(1 - a_1X - \cdots - a_tX^t\right)^{r+1}},\end{aligned}$$

where $P_r(X)$ is a polynomial of degree $< t(r+1)$. Hence

$$v(X) - w(X) = \frac{P(X)}{\left(1 - a_1X - \cdots - tX^t\right)^k},$$

where $P(X)$ is a polynomial of degree $< tk$. Hence $v_n(k) = w_n(k)\ \forall n \geqslant kt$, where

$$\sum_{n=0}^{\infty} w_n(k)X^n = \left(1 - a_1X - \cdots - a_tX^t\right)^k w(X).$$

Since $k!$ divides $D^{n-k}w_n$, we see that $k!$ divides $D^n w_n(k) = D^n v_n(k)$ as asserted.

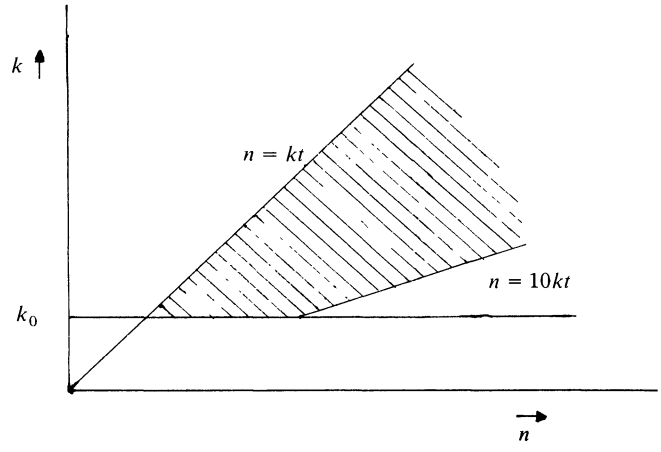
Proof of the Claim. It is sufficient to prove that $\sum_{n=0}^{\infty} v_n(k)X^n \in \mathbb{Q}[X]$ for some $k \in \mathbb{N}$. From the Lemma it follows that if $v_n(k) \neq 0$ and $n \geqslant kt$, then

$$k! \leqslant \left|D^n v_n(k)\right| \leqslant c_2(AD)^n C^k.$$

Hence, if $k! > c_2(AD)^n C^k$ and $n \geqslant kt$ then $v_n(k) = 0$. Choose k_0 so large that $k! > c_2(AD)^{10kt}C^k\ \forall k \geqslant k_0$. Then

$$v_n(k) = 0 \quad \text{for all } k \geqslant k_0,\ kt \leqslant n \leqslant 10kt. \tag{8}$$

This situation can be pictured as follows: If a point (n, k) falls in the shaded region, we have automatically $v_n(k) = 0$ according to (8). So, finally, by (7) and induction on $n - 10kt$, it follows that $v_n(k) = 0$ for all (n, k) in the infinite triangular region $k_0 \leqslant k \leqslant n/10t$. Thus we conclude that $v_n(k_0) = 0\ \forall n \geqslant k_0 t$ and hence $\sum_{n=0}^{\infty} v_n(k)X^n \in \mathbb{Q}[X]$, which proves our claim.



REFERENCES

1. J. P. Bézivin and P. Robba, Rational solution of linear differential operations, *J. Australian Math. Soc.* (to appear).
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, 1959.
3. D. Hilbert, Gesammelte Abhandlungen, vol. I, Chelsea, 1965, pp. 1–4.
4. A. Baker, Transcendental Number Theory, Cambridge Univ. Press, 1975.

Lucas's Theorem and Some Related Results for Extended Pascal Triangles

RICHARD C. BOLLINGER and CHARLES L. BURCHARD, *Penn State at Erie, Behrend College*

RICHARD BOLLINGER received his Ph.D. in 1961 from the University of Pittsburgh, with a thesis in Summability Theory under the direction of George Laush; he has since also worked in other areas of analysis, statistics, reliability theory, and, more recently, combinatorics. He has been at Penn State since 1962, most of that time at Penn State, Erie, The Behrend College, where he is currently Professor of Mathematics.



CHARLES BURCHARD did his undergraduate work in mathematics at Allegheny College and received an M.A. in mathematics and an M.S. in computer science from Penn State. He has been at Penn State, Erie, The Behrend College, since 1983, where he is currently Instructor of Computer Science.



1. Introduction. The result known as Lucas's Theorem first appeared in 1878 in *The American Journal of Mathematics* [**1**] (more accessible sources are [**2**], p. 271, items 76 and 77, and a complete translation in [**3**]), and reads as follows. Let p be a prime, and for integers n and k set $n = \Sigma n_i p^i$, $k = \Sigma k_i p^i$, with $0 \leqq n_i, k_i < p$; then

$$\binom{n}{k} \equiv \prod \binom{n_i}{k_i} \pmod{p}, \tag{1.1}$$

where we set $\binom{n_i}{k_i} = 0$ if $k_i > n_i$. Proofs other than Lucas's have been given (e.g., [**4**]), the congruence has been generalized in various ways (e.g., [**5**], [**6**]), and many related results have been elaborated (see [**7**] and its many references for an extensive discussion of this and related topics).

There seem to be, however, no corresponding developments for extended Pascal triangles T_m (as defined below), and in Section 2 we consider the extension of Lucas's congruence to the entries $C_m(n, k)$ of T_m; both statement and proof are straightforward, although the form is more awkward to use than (1.1). Section 3 then uses some results related to the extension to investigate, as in [**4**], [**7**], some divisibility properties of $C_m(n, k)$; in particular, when is $C_m(n, k) \not\equiv 0 \pmod{p}$ for k in the range $0 \leqq k \leqq (m-1)n$?

The extended Pascal triangles arise, by analogy with the ordinary Pascal triangle, as the (left-justified) arrays of the coefficients in the expansion of $(1 + x + x^2 + \cdots + x^{m-1})^n$. That is, the array T_m has in row n, column k, the number $C_m(n, k)$ defined for $m, n, k \geqq 0$ by

$$(1 + x + x^2 + \cdots + x^{m-1})^n = \sum_{k=0}^{(m-1)n} C_m(n, k) x^k. \tag{1.2}$$

Omitting calculational details, we list below some properties of these arrays for $m, n, k \geqq 0$.

a) The coefficients in T_m may be obtained from the relations

$$C_m(0, k) = \delta_{0k}$$

$$C_m(n+1, k) = \sum_{j=0}^{m-1} C_m(n, k-j).$$

b) It follows then that T_0 is the all-zero array except for $C_0(0,0) = 1$, and all the rows of T_1 consist of a one followed by zeros. For $m \geqq 2$, T_m is the array whose $n = 0$ row is a one followed by zeros, whose $n = 1$ row is m ones followed by zeros, and any of whose entries in subsequent rows is the sum of the m entries just above and to the left in the preceding row.

c) For $m > 0$ there are $(n(m-1)+1)$ nonzero entries in row n of T_m, and the symmetry relation among these is

$$C_m(n, k) = C_m(n, (m-1)n - k), \qquad 0 \leqq k \leqq (m-1)n.$$

d) T_2 is, of course, *the* Pascal triangle, and $C_2(n, k) = \binom{n}{k}$. The $C_m(n, k)$ can be expressed in terms of ordinary binomial coefficients,

$$C_m(n, k) = \sum_j (-1)^j \binom{n}{j} \binom{n - 1 + k - mj}{n - 1},$$

but for present purposes it is preferable to think of T_m as having an independent existence and a triangle of its "own," which could be tabulated as easily and familiarly as T_2.

e) As an example, here are the first few rows of T_3:

n \ k	0	1	2	3	4	5	6	7	8
0	1	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0
2	1	2	3	2	1	0	0	0	0
3	1	3	6	7	6	3	1	0	0
4	1	4	10	16	19	16	10	4	1

In the following sections, letters d, e, k, m, n, r, s, and so on, will denote nonnegative or positive integers as specified by the context, and p will denote a prime. We also adopt a modification of Singmaster's notation $N(n, p)$ [7] by defining $N_m(n, p)$ as the number of values of k for which $C_m(n, k) \not\equiv 0 \pmod{p}$.

2. The Extension of Lucas's Theorem. The proof given below is direct, and only depends on a simple congruence and the fact that the entries in the nth row of T_m are the coefficients in the expansion (1.2).

THEOREM 1. *Let p be a prime, and let*

$$n = n_0 + n_1 p + \cdots + n_r p^r, \qquad 0 \leqq n_i < p$$

$$k = k_0 + k_1 p + \cdots + k_t p^t, \qquad 0 \leqq k_i < p, \qquad 0 \leqq k \leqq (m-1)n. \quad (2.1)$$

Then

$$C_m(n, k) \equiv \sum_{(s_0, \ldots, s_r)} \prod_{i=0}^{r} C_m(n_i, s_i) \pmod{p} \quad (2.2)$$

where the sum is taken over all $(r+1)$*-tuples* $(s_0, s_1, \ldots, s_r)$ *such that i)* $s_0 + s_1 p + \cdots + s_r p^r = k$, *and ii)* $0 \leqq s_i \leqq (m-1)n_i$; *if* k *is not representable in this form then certainly* $C_m(n, k) \equiv 0 \pmod{p}$.

Proof.

$$\begin{aligned}
\sum_{k=0}^{(m-1)n} C_m(n,k)x^k &= \left(1 + x + \cdots + x^{m-1}\right)^n \\
&= \prod_{i=0}^{r} \left(1 + x + \cdots + x^{m-1}\right)^{n_i p^i} \\
&\equiv \prod_{i=0}^{r} \left(1 + x^{p^i} + x^{2p^i} + \cdots + x^{(m-1)p^i}\right)^{n_i} \pmod{p} \\
&= \prod_{i=0}^{r} \left\{ \sum_{s_i=0}^{(m-1)n_i} C_m(n_i, s_i) x^{s_i p^i} \right\} \\
&= \sum_{k=0}^{(m-1)n} \left\{ \sum_{(s_0,\ldots,s_r)} \prod_{i=0}^{r} C_m(n_i, s_i) \right\} x^k,
\end{aligned}$$

and the result follows by equating coefficients of x^k. All the steps here involve straightforward replacements, except the one involving the congruence in the third line, which may be proved by induction. Note that for $m = 2$, conditions (i) and (ii) imply that the expression of k in terms of $s_0, \ldots, s_r$ is just the unique base-p representation of $k (t = r,\ s_i = k_i,\ 0 \leqq k_i < p)$ and so there is only one term in the sum of products, which in this case reduces to the original form (1.1).

Example. $C_3(8, 4) \equiv 2 \pmod 3$, the calculation being as follows. Using (2.2), with $n = 8 = 22$, base 3 (so that $n_0 = 2,\ n_1 = 2$),

$$C_3(8,4) \equiv \sum_{(s_0, s_1)} \prod_{i=0}^{1} C_3(n_i, s_i) \pmod 3,$$

where the sum is taken over all (s_0, s_1) such that $s_0 + 3s_1 = 4$, and $0 \leqq s_0, s_1 \leqq 4$, which means here that (s_0, s_1) takes on the values $(1, 1)$ and $(4, 0)$. Thus,

$$\left.\begin{aligned}
C_3(8,4) &\equiv C_3(2,1) \cdot C_3(2,1) + C_3(2,4) \cdot C_3(2,0) \\
&= 2 \cdot 2 + 1 \cdot 1 \\
&\equiv 1 + 1 \\
&= 2
\end{aligned}\right\} \pmod 3.$$

Even for m, n small there may be many terms in a sum of products such as that above, because the conditions (i), (ii) may allow a given value of k to be represented in many ways.

It is also possible that the conditions do not allow of any representation of k of the specified form.

Example. $C_3(5, 7) \equiv 0 \pmod 2$ because $k = 7$ here is not representable in the required form. That is, with $n = 5 = 101$, base 2, we have $n_0 = 1,\ n_1 = 0,\ n_2 = 1$, so that $0 \leqq s_0 \leqq 2,\ s_1 = 0,\ 0 \leqq s_2 \leqq 2$, but with these restrictions $k = 7$ is not representable in the form $s_0 \cdot 1 + 0 \cdot 2 + s_1 \cdot 2^2$, which can only assume the values $0, 1, 2, 4, 5, 6, 8, 9, 10$.

It may be worthwhile to point out that even though the conditions (i), (ii) may not allow some values of k to be represented, as in the previous example, the full range of k is allowed for, since with $n = n_0 + n_1 p + \cdots + n_r p^r$, we always have

$$0 \leqq k = s_0 + s_1 p + \cdots + s_r p^r$$
$$\leqq (m-1)(n_0 + n_1 p + \cdots + n_r p^r) = (m-1)n.$$

The coefficients k_i in the strict p-ary representation of k really play no role in all this for $m > 2$, being superseded by the s_i.

3. Related Results. In this section, we take up the question of the number of values of k for which $C_m(n, k) \not\equiv 0 \pmod{p}$; the main results are exact determinations of $N_m(n, p)$ for the cases $m = p$ and $m = p^e$, $p \geqq 2$. Among the things one might want to know: what the minimal and maximal values of $N_m(n, p)$ are and where they occur; and a formula, if possible, for $N_m(n, p)$. These questions are already settled in the case of $m = 2$ ([7], results and references, sections 6, 7) and the main results are as follows.

1) With $n = n_0 + n_1 p + \cdots + n_r p^r$,

$$N_2(n, p) = \prod_{i=0}^{r} (n_i + 1). \tag{3.1.1}$$

2) $N_2(n, p) \geqq 2$, and $N_2(n, p) = 2$ if and only if $n = p^e$. (3.1.2)

3) $N_2(n, p) \leqq n + 1$, and $N_2(n, p) = n + 1$ if and only if $n = ap^s - 1$, with $1 \leqq a \leqq p$. (3.1.3)

These results might be derived by ad hoc arguments, but are usually obtained as consequences of Lucas's Theorem.

We briefly note that these results could also be argued from the point of view of the extension of Lucas's Theorem, since, for example, the number of valid $(r+1)$-tuples $(s_0, s_1, \ldots, s_r)$ occurring in (2.2) and satisfying conditions (i), (ii) is

$$\prod_{i=0}^{r} ((m-1)n_i + 1),$$

which for $m = 2$ gives the formula (3.1), and is in this case just the number of $C_2(n, k)$ not congruent to zero (mod p); the other results can then be derived from this.

To evaluate $N_p(n, p)$ we first note that when m is a prime, p, the polynomials $1 + x + x^2 + \cdots + x^{p-1}$ in (1.2), whose powers give rise to the corresponding Pascal-T triangles T_p, are just the cyclotomic polynomials of prime index, $\Phi_p(x)$ [**8**, p. 76 ff.] (we will also need the initial cyclotomic polynomial $\Phi_1(x) = x - 1$). Further, the fact that we want to consider congruences, mod p, puts us in the situation of working with these cyclotomic polynomials over a finite field of characteristic p, so that we have available the following results, usually used in factoring polynomials over finite fields.

THEOREM [**8**, p. 79]. *If p is a prime and* $\gcd(n, p) = 1$, *then for* $e \geqq 1$

$$\Phi_{np^e}(x) = \Phi_{np}(x^{p^{e-1}}) \tag{3.2}$$

$$\Phi_{np^e}(x) = (\Phi_n(x))^{p^e - p^{e-1}} \text{ in a field of characteristic } p. \tag{3.3}$$

Using (3.3) with $n = 1$, $e = 1$, we have that $\Phi_p(x) \equiv \Phi_1(x)^{p-1} = (x-1)^{p-1}$

(mod p), i.e.,

$$(1 + x + x^2 + \cdots + x^{p-1})^n \equiv (-1)^{(p-1)n}(1 - x)^{(p-1)n} \pmod{p}.$$

Now, neither the sign in front, nor the fact that we have powers of $(1 - x)$ rather than powers of $(1 + x)$, has any effect in this context on whether a given coefficient in the expansion is congruent to zero (mod p). What this means is that, since the expression on the right is binomial, we can use the formula (3.1.1), appropriately adjusted, to evaluate $N_p(n, p)$. That is, $N_p(n, p)$ is given by $N_2((p - 1)n, p)$. We summarize this discussion in the following theorem.

THEOREM 2. *Let the p-ary expansion of* $(p - 1)n$ *be* $a_0 + a_1 p + \cdots + a_r p^r$. *Then in* T_p *the number of entries* $C_p(n, k)$ *in line n not congruent to zero* (mod p) *is given by* (*cf.* 3.1.1)

$$N_p(n, p) = \prod_{i=0}^{r} (1 + a_i). \tag{3.4}$$

Some other aspects of $N_p(n, p)$, *which parallel known results for* $N_2(n, p)$, *are given in the following corollaries.*

COROLLARY 2.1. *For* $n \geqq 1$, *the number of entries not congruent to zero* (mod p) *in any line n of* T_p *is at least p, and equals p exactly when n is a power of p. That is* (*cf.* 3.1.2),

$$N_p(n, p) \geqq p, \quad \textit{and} \quad N_p(n, p) = p \quad \textit{iff } n = p^r.$$

Proof. We know that $N_p(n, p)$ is given by (3.4), with the a_i being the digits in the representation of $(p - 1)n$. But these digits must add up to a multiple of $(p - 1)$, and so in the expanded form of (3.4) we have

$$\prod_{i=0}^{r} (1 + a_i) = 1 + (a_0 + \cdots + a_r) + (\text{sums of products of the } a_i)$$

$$= 1 + s(p - 1) + (\text{nonnegative terms}),$$

from which we can see that we always have $N_p(n, p) \geqq p$, and that $N_p(n, p) = p$ if and only if one of the a_i is $(p - 1)$ and the rest are zero. But then the corresponding digits in the p-ary expansion of n itself are, respectively, 1 and the rest zeros, which means n is a power of p. Thus, $N_p(p, n)$ is minimal iff n is a power of p.

COROLLARY 2.2. *For* $n \geqq 1$, *the number of entries not congruent to zero* (mod p) *in any line n of* T_p *is at most* $(p - 1)n + 1$, *and equals* $(p - 1)n + 1$ *exactly when n is of the form* $(p^r - 1)/(p - 1)$. *That is* (*cf.* 3.1.3),

$$N_p(n, p) \leqq (p - 1)n + 1, \quad \textit{and}$$

$$N_p(p, n) = (p - 1)n + 1 \quad \textit{iff } n = (p^r - 1)/(p - 1).$$

Proof. The inequality is trivial, since there are only $(p - 1)n + 1$ entries in line n. For equality, we must have $N_p(n, p) = N_2((p - 1)n, p) = (p - 1)n + 1$. But then we know from [7, Th. 14] that the argument of N_2 must be of the form $ap^r - 1$, with $1 \leqq a < p$. That is, $(p - 1)n = ap^r - 1$, or $n = (ap^r - 1)/(p - 1)$. But n here is an integer, so that $a = 1$. Thus, $N_p(n, p)$ is maximal iff $n = (p^r - 1)/(p - 1)$.

COROLLARY 2.3. *For Pascal-T triangles T_m, where m is a power of a prime p, we have*

$$N_{p^e}(n, p) = N_p(n(p^e - 1)/(p - 1), p)$$
$$= N_2(n(p^e - 1), p).$$

Proof. Direct calculation using (3.2) shows that

$$(1 + x + x^2 + \cdots + x^{p^e-1}) \equiv (1 + x + x^2 + \cdots + x^{p-1})^{\binom{p^e-1}{p-1}} \pmod{p},$$

from which the result follows. Maximal and minimal values of N_{p^e} can then be obtained as in corollaries 2.1 and 2.2.

Example. For the rows of T_5 (mod 5), Theorem 2 and Corollaries 2.1, 2.2 give the following.

a) In row $n = 10$, the number of entries $C_5(10, k)$ not congruent to zero (mod 5) is $N_5(10, 5) = 8$ since the base-5 representation of $4 \cdot 10 = 40$ is 130, so that $n_0 = 0$, $n_1 = 3$, $n_2 = 1$, and (3.4) gives $(1 + 0)(1 + 3)(1 + 1)$.

b) For $n \geqq 1$, $N_5(n, 5) \geqq 5$, and $N_5(n, 5) = 5$ iff $n = 1, 5, 25, 125, \ldots$.

c) For $n \geqq 1$, $N_5(n, 5) \leqq 4n + 1$, and $N_5(n, 5) = 4n + 1$ iff $n = 1, 6, 31, 156, \ldots$, since these are the values of $(5^r - 1)/4$, $r = 1, 2, 3, \ldots$, the values whose base-5 representations are all 1's.

Lastly, we also have the following density result, where the density here is "triangular," as defined below; cf. also [9] in this connection.

COROLLARY 2.4. *In T_p, "almost all" entries $C_p(n, k)$ are divisible by p.*

Proof. Let m be an integer such that $p^e \leqq (p - 1)m \leqq p^{e+1}$. Then $\sum_{n=0}^{m} N_p(n, p)$ is the number of $C_p(n, k)$ not divisible by p, and lying in the "triangle" consisting of rows $n = 0$ through $n = m$ of T_p; the total number of elements in this triangle will be $s_m = (1/2)(m + 1)[(p - 1)m + 2]$. Now,

$$\sum_{n=0}^{m} N_p(n, p) = \sum_{n=0}^{m} N_2((p - 1)n, p) \leqq \sum_{n=0}^{(p-1)m} N_2(n, p)$$
$$\leqq p(p + 1)\left(\frac{p + 1}{2p}\right)^e ((p - 1)m)((p - 1)m + 1),$$

where the last inequality follows from Th. 1(b) of [9]. But then $(\sum_{n=0}^{m} N_p(n, p))/s_m$ goes to zero as m (and thus e) goes to infinity.

In conclusion, we remark that results corresponding to those of Theorem 2 and its corollaries are likely to be less straightforward for T_m (mod p), m composite, or even T_q (mod p), p, q distinct primes. It is true, for example in T_3 (mod 2), that $N_3(n, 2) \geqq 3$ for any $n \geqq 1$, since we always have $C_3(n, 0) = C_3(n, 2n) = 1$, and it may be shown by induction, using the recurrence relation for the coefficients, that $C_3(n, n) \equiv 1 \pmod{2}$. It is also true that $N_3(n, 2) = 3$ exactly when n is a power of 2. On the other hand, $N_3(n, 2)$ does not take on the value $2n + 1$ for any $n > 1$, as is shown by the following brief argument. The patterns of the first four elements in rows $n = 2$ to $n = 5$ of T_3 (mod 2) are 1010, 1101, 1000, and 1110. After that the cycle of patterns is repeated because the first four entries in a row are determined by

the first four entries of the preceding row; hence, every row contains at least one 0.

Even so, the present discussion extends much of what is known for T_2 to a whole range of "triangles", obtaining some fairly obvious results but also some which might not have been predicted from T_2, and leaving evident several interesting open questions.

The authors are indebted to the referee for a number of helpful suggestions.

REFERENCES

1. Edouard Lucas, Théorie des Functions Numériques Simplement Periodiques, *American J. Math.*, 1 (1878), 184–240, 289–321.
2. L. E. Dickson, History of the Theory of Numbers, vol. I, New York, Chelsea, 1952.
3. E. Lucas, The Theory of Simply Periodic Numerical Functions (translated by S. Kravitz, D. Lind), The Fibonacci Association, Santa Clara, 1969.
4. N. J. Fine, Binomial coefficients modulo a prime, *Amer. Mathem. Monthly*, 54 (1947) 589–592.
5. G. S. Kazandzidis, Congruences on the binomial coefficients, *Bull. Soc. Math. Grece (NS)*, 9 (1968) 1–12. MR 42, #182.
6. David Singmaster, Notes on binomial coefficients I, *J. London Math. Soc.*, (2) 8 (1974) 545–548.
7. _____, Divisibility of Binomial and Multinomial Coefficients by Primes and Prime Powers, *18th Anniversary Volume of the Fibonacci Association*, 1980, 98–113.
8. R. J. McEliece, Finite Fields for Computer Scientists and Engineers, Kluwer Academic Publishers, Boston, 1987.
9. David Singmaster, Notes on binomial coefficients III, *J. London Math. Soc.*, (2) 8 (1974) 555–560.

There are three kinds of mathematicians: those who can count, and those who can't.

(this is due to Marcia Sward, Solomon A. Garfunkel,
A. G. Howson, . . . , and possibly others)

The Remainder in Taylor's Formula*

ESTEBAN I. POFFALD, *Wabash College, Crawfordsville, IN*

ESTEBAN I. POFFALD completed his undergraduate studies at Universidad Catolica de Chile, Temuco, Chile. He received an M.S. in Mathematics from Universidad Tecnica de Chile for work done under the direction of G. Riera. His Ph.D. dissertation was done under the supervision of S. Reich at the University of Southern California. He has been at Wabash since 1985.



We present here a mean-value theorem that generalizes the Taylor-Lagrange formula. The result arises in a natural way when one studies the asymptotic behavior of the remainder term of the formula. As an *application* of our result, we derive several numerical schemes to approximate the solution to initial-valued first order differential equations.

To begin, let us recall the Taylor-Lagrange formula. For convenience, we work in intervals of the form $[0, x]$.

THEOREM. (Taylor-Lagrange formula). *If f is continuous in $[0, x]$, $f^{(n-1)}(0)$ exists and $f^{(n)}(t)$ exists in $(0, x)$, then there exists a ξ in $(0, x)$ such that*

$$f(x) = p_{n-1}(x) + f^{(n)}(\xi)\frac{x^n}{n!}, \tag{1}$$

where

$$p_{n-1}(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2!} + \cdots + f^{(n-1)}(0)\frac{x^{n-1}}{(n-1)!}$$

is the Taylor polynomial of order $n-1$ for f about 0.

Our first result concerns the asymptotic behavior as $x \to 0^+$ of the number ξ in the theorem above:

THEOREM 1. *With notation as in the previous theorem, if $f^{(n+1)}(t)$ exists in $[0, x]$, is continuous from the right at $t = 0$ and if $f^{(n+1)}(0) \neq 0$, then*

$$\lim_{x \to 0+} \frac{\xi}{x} = \frac{1}{n+1}.$$

Remark. The assumptions in this theorem imply that the number ξ is uniquely determined for x small enough.

*This work was supported in part by the Byron K. Trippet Research Stipend at Wabash College.

Proof of the theorem. In (1), we apply the Mean-Value theorem to $f^{(n)}(\xi)$: there is a number τ in $(0, \xi)$ such that

$$\begin{aligned} f(x) &= p_{n-1}(x) + f^{(n)}(\xi)\frac{x^n}{n!} \\ &= p_{n-1}(x) + \left[f^{(n)}(0) + f^{(n+1)}(\tau)\xi\right]\frac{x^n}{n!} \\ &= p_n(x) + f^{(n+1)}(\tau)\xi\frac{x^n}{n!}. \end{aligned}$$

On the other hand, by the Taylor-Lagrange formula

$$f(x) = p_n(x) + f^{(n)}(\sigma)\frac{x^{n+1}}{(n+1)!} \quad \text{for some } \sigma \text{ in } (0, x).$$

Therefore,

$$f^{(n)}(\tau)\xi = f^{(n)}(\sigma)\frac{x}{n+1}.$$

Since $f^{(n+1)}$ is right-continuous and nonzero at $t = 0$, the conclusion follows.

With notation as above, $f^{(n)}(\xi)\frac{x^n}{n!}$ can be viewed as the error made in approximating $f(x)$ by $p_{n-1}(x)$. In view of Theorem 1, one would expect that replacing ξ by $x/(n+1)$ and approximating $f(x)$ by

$$A_0(x) = p_{n-1}(x) + f^{(n)}\left(\frac{x}{n+1}\right)\frac{x^n}{n!}$$

would result in an approximation to $f(x)$ of order at least that of $p_n(x)$. In fact, it turns out that this approximation is of the same order as $p_{n+1}(x)$. We no longer require that $f^{(n+1)}(0) \neq 0$.

THEOREM 2. *If $f^{(n+2)}(t)$ exists and is continuous in $[0, x]$, then there exists a ξ in $(0, x)$ such that*

$$\begin{aligned} f(x) &= A_0(x) + \frac{n}{2(n+1)}f^{(n+2)}(\xi)\frac{x^{n+2}}{(n+2)!} \\ &= p_{n-1}(x) + f^{(n)}\left(\frac{x}{n+1}\right)\frac{x^n}{n!} + \frac{n}{2(n+1)}f^{(n+2)}(\xi)\frac{x^{n+2}}{(n+2)!}. \end{aligned}$$

Proof. Using Taylor's formula with integral remainder, we can write

$$f^{(n)}\left(\frac{x}{n+1}\right) = f^{(n)}(0) + f^{(n+1)}(0)\frac{x}{n+1} + \int_0^{\frac{x}{n+1}} f^{(n+2)}(t)\left[\frac{x}{n+1} - t\right]dt.$$

Therefore,

$$A_0(x) = p_{n+1}(x) + \frac{x^n}{n!}\int_0^{\frac{x}{n+1}} f^{(n+2)}(t)\left[\frac{x}{n+1} - t\right]dt$$

but

$$f(x) = p_{n+1}(x) + \frac{1}{(n+1)!}\int_0^x f^{(n+2)}(t)(x-t)^{n+1}\,dt.$$

It follows that

$$\begin{aligned} f(x) - A_0(x) &= \frac{1}{(n+1)!}\int_0^x f^{(n+2)}(t)(x-t)^{n+1}\,dt \\ &\quad - \frac{x^n}{n!}\int_0^{\frac{x}{n+1}} f^{(n+2)}(t)\left[\frac{x}{n+1} - t\right]dt \\ &= \int_0^{\frac{x}{n+1}} f^{(n+2)}(t)\left[\frac{(x-t)^{n+1}}{(n+1)!} - \frac{x^n}{n!}\left[\frac{x}{n+1} - t\right]\right]dt \\ &\quad + \frac{1}{(n+1)!}\int_{\frac{x}{n+1}}^x f^{(n+2)}(t)(x-t)^{n+1}\,dt. \end{aligned}$$

Now, let

$$g(t) = \frac{(x-t)^{n+1}}{n+1} - x^n\left(\frac{x}{n+1} - t\right).$$

Then $g(0) = 0$ and $g'(t) > 0$ in $(0, x]$, so that $g(t) \geqslant 0$ in that interval. Therefore, we can apply the Mean-Value theorem for integrals to deduce that there are numbers ξ_1 and ξ_2 in $(0, x)$ such that

$$\begin{aligned} f(x) - A_0(x) &= \frac{1}{n!}f^{(n+2)}(\xi_1)\int_0^{\frac{x}{n+1}} g(t)\,dt \\ &\quad + \frac{1}{(n+1)!}f^{(n+2)}(\xi_2)\int_{\frac{x}{n+1}}^x (x-t)^{n+1}\,dt. \end{aligned}$$

Finally, an application of the Intermediate Value Theorem concludes the proof.

In view of the way in which the statement of Theorem 2 follows naturally from Theorem 1, it should be clear at this point that we should try to determine the asymptotic behavior as $x \to 0^+$ of the number ξ in Theorem 2 and then establish a result similar to the one just presented. We are thus led to consider approximations to $f(x)$ of the form

$$\begin{aligned} A_k(x) &= p_{n-1}(x) + M_0 f^{(n)}(c_0 x)\frac{x^n}{n!} + M_2 f^{(n+2)}(c_2 x)\frac{x^{n+2}}{(n+2)!} \\ &\quad + \cdots + M_{2k} f^{(n+2k)}(c_{2k}x)\frac{x^{n+2k}}{(n+2k)!}, \end{aligned}$$

where for a given n, the coefficients M_{2i} and c_{2i} depend only on i. Clearly, we have $M_0 = 1$ and $c_0 = \dfrac{1}{n+1}$.

In the spirit of Theorem 2, we want $A_k(x)$ to be an approximation to $f(x)$ of order $n + 2k + 2$. By expanding $f^{(n+2j)}(c_{2j}x)$ in Taylor polynomials about 0 and

collecting terms, we see that the sequences $\{M_{2j}\}$ and $\{c_{2j}\}$ can be obtained recursively from

$$\frac{M_0}{n!}\frac{c_0^{2k}}{(2k)!} + \frac{M_2}{(n+2)!}\frac{c_2^{2k-2}}{(2k-2)!} + \cdots + \frac{M_{2k}}{(n+2k)!} = \frac{1}{(n+2k)!} \tag{2}$$

$$\frac{M_0}{n!}\frac{c_0^{2k+1}}{(2k+1)!} + \frac{M_2}{(n+2)!}\frac{c_2^{2k-1}}{(2k-1)!} + \cdots + \frac{M_{2k}}{(n+2k)!}\frac{c_{2k}}{1!} = \frac{1}{(n+2k+1)!}. \tag{3}$$

Then, one verifies readily that if $f^{(n+2k+2)}$ exists and is continuous at 0,

$$\lim_{x\to 0}\frac{f(x) - A_k(x)}{x^{n+2k+2}} = \frac{M_{2k+2}}{(n+2k+2)!}f^{(n+2k+2)}(0),$$

and, therefore, $A_k(x)$ is indeed an approximation to $f(x)$ of order $n + 2k + 2$. In fact, the following mean-value theorem holds.

THEOREM 3. *If $f^{(n+2k)}(t)$ exists and is continuous in $[0, x]$, then there exists a number ξ in $(0, x)$ such that*

$$f(x) = A_{k-1}(x) + M_{2k}f^{(n+2k)}(\xi)\frac{x^{n+2k}}{(n+2k)!}. \tag{4}$$

Moreover, $0 < M_{2j} < 1$ for $j \geqslant 1$ and $0 < c_{2j} < 1$ for $j \geqslant 0$.

Proof. By induction on k. For $k = 1$, this is just Theorem 2. Let us assume that the result holds for $k > 1$. Assume also that $0 < M_{2j} < 1$ for $j = 1, \ldots, k$ and $0 < c_{2j} < 1$ for $j = 0, \ldots, k-1$. Finally, as part of the induction hypothesis, assume that $g_{k-1}(\lambda) > 0$ for $0 < \lambda < 1$, where

$$g_k(\lambda) = \frac{(1-\lambda)^{n+2k+1}}{(n+2k+1)!} - \frac{M_0(c_0-\lambda)_+^{2k+1}}{n!(2k+1)!} - \cdots - \frac{M_{2k}(c_{2k}-\lambda)_+}{(n+2k)!1!}.$$

As usual, z_+ denotes the largest of 0 and z, for any real number z. (That $g_0(\lambda) > 0$ for $0 < \lambda < 1$ follows from Bernoulli's inequality.) First of all, observe that if f has $n + 2k + 1$ continuous derivatives and $f^{(n+2k+1)}(0) \neq 0$, then $\lim_{x\to 0+}(\xi/x) = c_{2k}$, where ξ in $(0, x)$ satisfies

$$f(x) = A_{k-1}(x) + \frac{M_{2k}}{(n+2k)!}f^{(n+2k)}(\xi).$$

Therefore, $0 \leqslant c_{2k} \leqslant 1$. This, together with the induction hypotheses, implies that $g_k(0) = g_k(1) = 0$. Notice also that $g_k''(\lambda)$ exists for $\lambda \neq c_{2k}$ and $g_k''(\lambda) = g_{k-1}(\lambda)$. This implies that g_k is convex in $[0, c_{2k}]$ and in $[c_{2k}, 1]$.

If $c_{2k} \neq 0, 1$, then $g_k''(\lambda) > 0$ for $\lambda \neq 0, c_{2k}, 1$ implies that g_k' is increasing in $[0, c_{2k}]$ and in $[c_{2k}, 1]$. We also have that $g_k'(0) = g_k'(1) = 0$, which implies that g_k is convex and increasing in $[0, c_{2k}]$ and it is convex and decreasing in $[c_{2k}, 1]$. The conclusion is that $g_k(\lambda) \geqslant 0$ in $[0, 1]$ with equality only for $\lambda = 0$ and 1.

If c_{2k} were 0 or 1, then we could carry out the same analysis as above, using one-sided derivatives instead. But $c_{2k} = 0$ or 1 would imply that $g_k \equiv 0$ in $[0, 1]$, which would force $g_{k-1} \equiv 0$ in $[0, 1]$, contradicting the induction hypotheses.

Now we are ready to complete the proof:
If f has $n + 2k + 2$ continuous derivatives in $[0, x]$, we have

$$f(x) - A_k(x) = f(x) - \sum_{j=0}^{k} M_{2j} f^{(n+2j)}(c_{2j}x) \frac{x^{n+2j}}{(n+2j)!}.$$

Now, expand each term in the right-hand side into its Taylor polynomial of degree $2k - 2j + 1$ about 0, with integral remainder, and collect terms. One gets

$$f(x) - A_k(x) = \int_0^x f^{(n+2k+2)}(t) \frac{(x-t)^{n+2k+1}}{(n+2k+1)!}\, dt - \sum_{j=0}^{x} \int_0^{c_{2j}x} M_{2j} f^{(n+2k+2)}(t) \frac{(c_{2j}x - t)^{2k-2j+1}}{(2k-2j+1)!} \frac{x^{n+2j}}{(n+2j)!}\, dt.$$

In each of the integrals above, substitute $t = \lambda x$. After rearranging, this gives

$$f(x) - A_k(x) = x^{n+2k+2} \int_0^1 f^{(n+2k+2)}(\lambda x) g_k(\lambda)\, d\lambda.$$

Since $g_k(\lambda) \geqslant 0$ in $[0, 1]$, we can apply the Mean-Value theorem for integrals to conclude the proof.

Remarks.

1. *From the point of view of applications*, it would be interesting to have a closed formula for the sequences $\{c_{2j}\}$ and $\{M_{2j}\}$ given by (2) and (3). *Numerical computations* suggest that for fixed n the sequence $\{c_{2j}\}$ is monotonically increasing while $\{M_{2j}\}$ decreases to 0.

2. If f is analytic in a disk with center at the origin and radius $R > 0$, then one can use the Cauchy-Hadamard estimates to deduce that the expansion

$$f(x) = p_{n-1}(x) + \sum_{j=0}^{\infty} M_{2j} f^{(n+2j)}(c_{2j}x) \frac{x^{n+2j}}{(n+2j)!}$$

is valid in the disk $|x| < R/2$. In practice, however, the region of convergence of this expansion appears to be at least as large as that of the power series expansion of f. (*Cf. Remark* 4 *below.*) The validity of this statement may depend on the behavior of the sequences M_{2i} and c_{2i} as $i \to \infty$.

3. For $n = 1$, both the statement and proof of our main result are simpler: elementary combinatorial identities reveal that in this case $c_{2j} = 1/2$ and $M_{2j} = 1/4^j$ satisfy (2) and (3). This suggests the following proof for $n = 1$: In the Taylor-Lagrange formula

$$f(b) = f(a) + f'(a)(b-a) + \cdots + f^{(m)}(a) \frac{(b-a)^m}{m!} + f^{(m+1)}(\xi) \frac{(b-a)^{m+1}}{(m+1)!},$$

take $m = 2k$. Put $a = x/2$ and $b = x$; then, put $a = x/2$ and $b = 0$, subtract and appeal to the Intermediate Value Theorem, to deduce that there is a ξ in $(0, x)$ such

that

$$f(x) = f(0) + \sum_{j=0}^{2k} \frac{2}{(2j+1)!} f^{(2j+1)}\left(\frac{x}{2}\right)\left[\frac{x}{2}\right]^{2j+1} + \frac{2}{(2k+1)!} f^{(2k+1)}(\xi)\left[\frac{x}{2}\right]^{2k+1}.$$

It follows that if

$$\lim_{k\to\infty} \frac{f^{(2k+1)}(\xi)}{(2k+1)!}\left[\frac{x}{2}\right]^{2k+1} = 0,$$

then

$$f(x) = f(0) + 2\sum_{j=0}^{\infty} \frac{1}{(2j+1)!} f^{(2j+1)}\left(\frac{x}{2}\right)\left[\frac{x}{2}\right]^{2j+1}. \tag{5}$$

4. We are grateful to W. Swift for calling our attention to the fact that if f is analytic at 0, then (5) is valid for all x in the interior of the Borel polygon of summability of f. Recall that the Borel polygon of f is constructed as follows: Let P be a point in the complex plane where f has a singularity and let O denote the origin. Draw the line Δ_P perpendicular to $\overline{OP}$ at P. Then, a point x is in the Borel polygon of summability if and only if x is in the half plane determined by Δ_P and 0, for all the singular points P of f. It then follows that x is in the interior of the Borel polygon if and only if f is analytic at every point of the closed disk with diameter $\overline{Ox}$. Therefore, if x belongs to the Borel polygon of f, then

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}\left(\frac{x}{2}\right)\left[z - \frac{x}{2}\right]^n$$

holds for all z such that $|z - (x/2)| \leqslant |x/2|$. As above, putting $z = x$ and then $z = 0$ and subtracting gives the result.

Example. If $f(x) = \ln(1 + x)$, then

$$f^{(2k+1)}(x) = \frac{(2k)!}{(1+x)^{2k+1}}.$$

Applying (5), we obtain

$$\ln(1+x) = 2\sum_{n=0}^{\infty} \frac{1}{2n+1}\left[\frac{x}{2+x}\right]^{2n+1}.$$

The Borel polygon of f is the half plane $\operatorname{Re}(x) > -1$. Therefore, this expansion is valid for all x with $\operatorname{Re}(x) > -1$. This should be contrasted with the circle of convergence of the Taylor series expansion for $f(x)$ about $x = 0$. Even in the common region of convergence, this expansion converges faster. Since

$$\arctan x = \frac{1}{2i}[\ln(1+ix) - \ln(1-ix)],$$

we also obtain

$$\arctan x = -i\sum_{n=0}^{\infty} \frac{1}{2n+1}\left[(x+2i)^{2n+1} - (x-2i)^{2n+1}\right]\left[\frac{x}{4+x^2}\right]^{2n+1},$$

which is valid for $-1 < \mathrm{Im}(x) < 1$. In particular, for all x real, $\arctan x$ may be computed using

$$\arctan x = \sum_{n=0}^{\infty} \frac{1}{2n+1} A_n(x)$$

where

$$A_0(x) = \frac{4x}{4+x^2}, \qquad A_1(x) = 4\left[\frac{x}{4+x^2}\right]^3 (3x^2 - 4)$$

and

$$A_{n+2}(x) = \left[\frac{x}{4+x^2}\right]^2 \left[2(x^2-4)A_{n+1}(x) - x^2 A_n(x)\right].$$

For example,

$$\arctan 2 = \sum_{k=0} \frac{(-1)^k}{4^k}\left[\frac{1}{4k+1} + \frac{1}{8k+6}\right]$$

and

$$\pi = \frac{16}{5} \sum_{n=0}^{\infty} \sum_{j=0}^{n} \frac{1}{2n+1} \binom{2n+1}{2j+1} \frac{(-4)^j}{5^{2n}}.$$

An Application. We will now use Theorem 3 to derive several numerical methods of approximating the solution of the initial-valued differential equation

$$y' = f(x, y) \qquad y(x_0) = y_0,$$

where f (and, therefore, y) is a sufficiently smooth function.

Fix $h > 0$. For each nonnegative integer j, we are seeking an approximation y_{j+1} to the exact value of the solution y at $x_{j+1} = x_0 + (j+1)h$. Let $f_j = f(x_j, y_j)$ and denote by E the error $y(x_{j+1}) - y_{j+1}$, assuming that $y(x_i) = y_i$ for $0 \leqslant i \leqslant j$ (E is the local discretization error). We will omit the derivation of the error term. Note, however, that in each case the error term can be derived using the main ideas in this note, namely expansion in Taylor polynomials with integral remainder followed by an application of the Mean-Value Theorem for integrals.

(i). The Modified Euler's Method: for $n = 1$ and $k = 1$, we apply (4) (in the interval $[x_{j-1}, x_{j+1}]$) to obtain the approximation $y_{j+1} = y_{j-1} + 2hf_j$ with $E = y'''(\xi)(h^3/3)$.

(ii). For $n = 2$ and $k = 1$ and the interval $[x_{j-2}, x_{j+1}]$, we get

$$y(x_{j+1}) \approx y(x_{j-2}) + 3hy'(x_{j-2}) + \frac{9h^2}{2} y''(x_{j-1}).$$

Now, using the approximation $y''(x_{j-1}) \approx [y'(x_j) - y'(x_{j-2})]/2h$ and rearranging, we get $y_{j+1} = y_{j-2} + (3h/4)[f_{j-2} + 3f_j]$ with the error term $E = (3/4)y^{(4)}(\xi)h^4$.

This approximation does not appear to be one of the standard multistep methods of elementary numerical analysis. However, applying it to approximate $\int_a^b g(t)\,dt$,

we obtain

$$\int_a^b g(t)\,dt \approx \frac{1}{4}(b-a)\left[g(a) + 3g\left(\frac{a+2b}{3}\right)\right].$$

This is clearly asymmetric with respect to a and b, which suggests also considering the approximation

$$\int_a^b g(t)\,dt \approx \frac{1}{4}(b-a)\left[g(b) + 3g\left(\frac{b+2a}{3}\right)\right]$$

and then averaging these to obtain

$$\int_a^b g(t)\,dt \approx \frac{1}{8}(b-a)\left[g(a) + 3g\left(\frac{a+2b}{3}\right) + 3g\left(\frac{2a+b}{3}\right) + g(b)\right],$$

which is the well-known Simpson's 3/8 Rule, with error $-(1/6480)(b-a)^5 g^{(4)}(\xi)$.

(iii). In (ii) above, we could equally well use the approximation $y''(x_{j-1}) \approx [y(x_j) - 2y(x_{j-1}) + y(x_{j-2})]/h^2$, which gives

$$y_{j+1} = \frac{1}{2}\left[11y_{j-2} - 18y_{j-1} + 9y_j\right] + 3hf_{j-2},$$

with $E = (3/4)y^{(4)}(\xi)h^4$.

(iv). Now, we use (4) with $n = 1$, $k = 2$ and the interval $[x_{j-3}, x_{j+1}]$ to get

$$y(x_{j+1}) \approx y(x_{j-3}) + y'(x_{j-1})4h + y'''(x_{j-1})\frac{(4h)^3}{24}.$$

Approximating $y'''(x_{j-1})$ as we approximated $y''(x_{j-1})$ in (iii) above, we derive the multistep method

$$y_{j+1} = y_{j-3} + \frac{4}{3}h\left[2f_{j-2} - f_{j-1} + 2f_j\right],$$

with the error $E = (14/45)y^{(5)}(\xi)h^5$.

(v). Again, we take $n = 1$ and $k = 2$, but we use the interval $[x_{j-1}, x_{j+1}]$. This gives $y(x_{j+1}) \approx y(x_{j-1}) + y'(x_j)2h + y'''(x_j)(2h)^3/24$. Using $y'''(x_j) \approx [y'(x_{j+1}) - 2y'(x_j) + y'(x_{j-1})]/h^2$, we obtain

$$y_{j+1} = y_{j-1} + (h/3)\left[f_{j+1} + 4f_j + f_{j-1}\right],$$

with $E = (-1/90)y^{(5)}(\xi)h^5$. Observe that this last formula, when interpreted as a numerical integration method, gives the well known Simpson's Rule. Also, notice that the formula is implicit, since y_{j+1} appears also in the right-hand side. However, using this formula in conjunction with the one derived in (iv), we obtain a well-known predictor-corrector method (Milne's method):

$$y_{j+1}^{(p)} = y_{j-3} + \frac{4h}{3}\left[2f_j - f_{j-1} + 2f_{j-2}\right]$$

$$y_{j+1} = y_{j-1} + \frac{h}{3}\left[f_{j+1}^{(p)} + 4f_j + f_{j-1}\right],$$

where $f_{j+1}^{(p)} = f(x_{j+1}, y_{j+1}^{(p)})$.

Conclusion. All the results above arose from studying the error term in the Taylor's polynomial approximation to a given function. By no means is this

restricted to Taylor's polynomials: in general, suppose that $A(f)$ is a numerical approximation to a functional $F(f)$ such that $F(f) = A(f) + E(f)$ for all f in an appropriate class of functions. If anything can be said about the error term $E(f)$, then perhaps one should try to use $A(f) + \bar{E}(f)$ as a better approximation to $F(f)$. Here, $\bar{E}(f)$ is an "estimate" of the error that does not depend on the particular function f. To illustrate this, consider the trapezoidal rule in numerical integration:

$$\int_a^b f(t)\,dt \approx \frac{(b-a)}{2}[f(a) + f(b)].$$

Now, the error of this approximation is $-f''(\xi)(b-a)^3/12$, for some ξ in (a, b). It can then be shown that if $f^{(3)}$ is continuous and nonzero at a, then, as $b \to a^+$, c approaches the midpoint of $[a, b]$, that is,

$$\lim_{b \to a+} \frac{(c-a)}{(b-a)} = \frac{1}{2}.$$

Therefore, we obtain the numerical integration method:

$$\int_a^b f(t)\,dt \approx \frac{(b-a)}{2}[f(a) + f(b)] - f''\left(\frac{a+b}{2}\right)\frac{(b-a)^3}{12}$$

with error $-(1/480)f^{(4)}(\xi)(b-a)^5$. Finally, we note that if we approximate

$$f''\left(\frac{a+b}{2}\right) \quad \text{by} \quad \left[f(a) - 2f\left(\frac{a+b}{2}\right) + f(b)\right] \Big/ \frac{(b-a)^2}{2}$$

and replace, we obtain the approximation

$$\int_a^b f(t)\,dt \approx \frac{(b-a)}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right],$$

i.e., Simpson's rule once again.

REFERENCES

1. R. Ellis and D. Gulick, Calculus with Analytic Geometry, 3rd ed., H. B. J. Publishers, 1986.
2. J. M. H. Olmsted, Real Variables, Appleton-Century-Crofts, Inc., 1959.
3. S. D. Conte and C. de Boor, Elementary Numerical Analysis, 2nd ed., McGraw-Hill, 1972.
4. J. Ll. Morris, Computational Methods in Elementary Numerical Analysis, John Wiley & Sons, 1983.
5. E. Borel, Leçons sur les Séries Divergentes, Gauthier-Villars, 1901.

LETTERS TO THE EDITOR

Editor:

A search for a larger "largest known prime number" was conducted at the Amdahl Benchmark Center in Santa Clara, California. The search consisted of selecting over 350,000 prime candidates of the form $h2^n - 1$ with $n > 216091$, sieving them for prime factors less than one trillion and running a Lucasian test [**1**] on over 7000 remaining candidates.

After over a year of searching for a prime number in this range, the prime

$$(391581) \times 2^{216,193} - 1$$

was found on August 6, 1989. The Lucasian primality tests take 33 minutes for numbers of this size on the Amdahl 1200E. The restriction $n > 216091$ was chosen so that any prime found would be the largest known, exceeding those discovered by Slowinski [**2**].

We wish to thank Jeff Young for confirming our result.

REFERENCES

1. H. Riesel, "Lucasian Criteria for the Primality of $N = h2^n - 1$," *Math. Comp.*, v. 23, 1969, pp. 869–875.
2. P. Ribenboim, The Book of Prime Number Records, Springer-Verlag, New York, 1988.

John Brown, Landon Curt Noll, B. K. Parady,
Joel F. Smith, and Sergio E. Zarantonello
Amdahl Corporation (M/S 265)
1250 East Arques Av.
Sunnyvale, CA 94088

and

Gene Ward Smith
Department of Mathematics
University of California
Berkeley, CA 94720

Editor:

In the May issue of the MONTHLY, George E. Andrews and R. J. Baxter present "A Motivated Proof of the Rogers-Ramanujan Identities". It seems to me that they missed a chance to give a simpler and better motivated approach to the proof.

For each positive integer i, let $F_i(q)$ be the generating function for partitions with parts at least i and differences between parts at least 2. Then it is clear that

$$F_i(q) = 1 + O(q^i)$$

and

$$F_i(q) - F_{i+1}(q) = q^i F_{i+2}(q)$$

for all i. Furthermore, it is easily seen that these equations determine all of the series $F_i(q)$. Thus if $G_1(q)$ is the generating function for partitions with parts of the forms $5m + 1$ and $5m + 4$, and $G_2(q)$ is the generating function for partitions with parts of the forms $5m + 2$ and $5m + 3$, and we put

$$G_{i+2}(q) = [G_i(q) - G_{i+1}(q)]/q^i$$

for all positive integers i, then to prove the two Rogers-Ramanujan identities, we need only show that

$$G_i(q) = 1 + O(q^i)$$

for all i. The authors came to this conclusion in a less direct way.

Raphael M. Robinson
University of California
Berkeley, CA 94720

Editor:

Indeed, if one starts out with *both* sides of the Rogers-Ramanujan identities, then one can effectively skip Sections 1, 2 and 4 of our paper. Robinson's comments are essentially equivalent to those in our first paragraph of Section 5. However, starting with both sides begs the question of motivation.

Our object was to start with the *one* side of the Rogers-Ramanujan identities concerning partitions into congruence classes mod 5. From there we examine a problem intrinsic to that side (Ehrenpreis's question), and we are consequently forced to consider partitions with differences at least 2 between parts. Or as we state in our introduction, "In answering this question we are led naturally to the Rogers-Ramanujan identities themselves."

George E. Andrews
Pennsylvania State University
University Park, PA 16802

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

A Link Between the Jordan Curve Theorem and the Kuratowski Planarity Criterion

CARSTEN THOMASSEN
Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104
and *Mathematical Institute, Bld. 303, Technical University of Denmark, 2800 Lyngby, Denmark*

1. Introduction. The purpose of this note is to show that, under certain mild conditions, a topological space has a Jordan Curve Theorem if and only if $K_{3,3}$ does not embed in the space. This sheds some light on exactly what properties of the plane the Jordan Curve Theorem uses.

The Euclidean plane has the following two fundamental properties:

(1) THE JORDAN CURVE THEOREM. *Every closed simple curve in the Euclidean plane partitions the plane into two connected components.*

(2) KURATOWSKI'S THEOREM. *A graph G can be embedded in the Euclidean plane if and only if G contains no subdivision of either of the graphs K_5 or $K_{3,3}$.*

All known proofs of the easy part of (2) (that K_5 and $K_{3,3}$ cannot be embedded in the plane) are based on (1). [**1**] contains a short proof of the easy part of (2) based only on the restriction of (1) to polygonal curves. Conversely, it is shown in [**2**] how (1) is an easy consequence of the easy part of (2). Thus there is a close relation between the two theorems, and H. Wilf raised the question if this relationship holds in other topological spaces as well. Here we shall answer the question in the affirmative for those topological spaces that cannot be separated by a simple arc. For spaces of low connectivity the relationship disappears. Every graph (which is defined in the next section) may be thought of as a topological space where the edges are simple arcs. Then a nonseparating, simple closed curve is the same as what in [**3**] is called a nonseparating induced cycle. The investigations in [**3**] show that the existence of such cycles depends more on edge density than planarity (or nonplanarity). One basic result in [**3**] says that every finite connected graph of minimum valency $\geqslant 3$ has a nonseparating induced cycle. For infinite graphs not much is known about nonseparating induced cycles.

2. Terminology. A *simple arc* A in a topological space X is the image of a continuous 1-1 function f: $[0,1] \to X$. If $0 \leqslant a < b \leqslant 1$, then $f([a, b])$ is the *segment* of A from $f(a)$ to $f(b)$. A *simple closed curve* is defined analogously except that now $f(0) = f(1)$. X is *arcwise connected* if any two elements are joined by a simple arc. A set $Y \subseteq X$ *separates* X if $X \setminus Y$ is not arcwise connected. We shall repeatedly make use of the following simple fact: If F_1, F_2 are disjoint closed sets in X and A is an arc from an element of F_1 to an element of F_2, then A contains a segment A' from an element of F_1 to an element of F_2 such that A' has no intermediate point in common with $F_1 \cup F_2$. Furthermore, the union of a finite number of arcs is compact and hence closed in case X is a Hausdorff space.

A *graph* is a (finite) set $V(G)$, called *vertices*, together with a set $E(G)$ of unordered pairs of vertices called *edges*. If the edge xy is present, we say that x is joined to y. The number of vertices joined to x is the *valency* of x. The *complete graph* K_n is the graph with n vertices and all possible edges. The *complete bipartite graph* $K_{m,p}$ is the graph whose vertex set can be partitioned into sets A and B with m, respectively p, vertices such that the edges are all mp edges joining A and B. A graph G is *embedded* in a topological space X if the vertices of G are distinct elements of X and every edge in G is a simple arc connecting in X the two vertices which it joins in G. Moreover, no two edges intersect except at a common vertex.

3. The link between curve separation and graph planarity. A special case of Euler's formula for graphs in the Euclidean plane says that no simple arc separates the plane. (A short proof is given in [**2**].) The following result holds for spaces with that connectivity property.

THEOREM. *Let X be a Hausdorff space which is arcwise connected and which is not separated by any simple arc. Assume also that X is not homeomorphic to a simple closed curve. Then the following statements are equivalent*:

(a) *Every simple closed curve separates X.*

(b) *Every simple closed curve separates X into precisely two arcwise connected components.*

(c) *$K_{3,3}$ cannot be embedded in X.*

(d) *Neither $K_{3,3}$ nor K_5 can be embedded in X.*

Proof. Clearly (b) $\Rightarrow$ (a) and (d) $\Rightarrow$ (c). It is also easy to prove that (c) $\Rightarrow$ (d). For suppose that K_5 is embedded in X. Let $v_1, v_2, \ldots, v_5$ be the vertices of K_5. Let p_1, p_2 be points on the edge v_1v_2 such that v_1, p_1, p_2, v_2 are distinct. Now consider an arc A in K_5 which connects p_1 and p_2 and which contains all of v_1, v_2, v_3, v_4, v_5. Since $X \setminus A$ is arcwise connected it contains an arc B from a point q_1 on the edge v_1v_2 to a point q_2 on another edge such that $B \cap K_5 = \{q_1, q_2\}$. It is now easy to find a $K_{3,3}$ in $K_5 \cup B$.

We shall now assume (c) and prove (b). Let C be any simple closed curve in X. Since X is not homeomorphic to a simple closed curve, there is a point $p \in X \setminus C$. Let A be a simple arc from p to C such that $A \cap C$ consists of one point q. Let A_1 be a segment of A from q to a point $p_1 \neq p$. Since $X \setminus A_1$ is arcwise connected it contains an arc A_2 from $A \setminus A_1$ to C having only its ends in common with $A \cup C$. Now $A \cup A_2$ contains a simple arc A_3 connecting q with another point q' on C such that $A_3 \cap C = \{q, q'\}$. Let A_4, A_5 be the two segments of C from q to q'. Since $X \setminus A_3$ is arcwise connected, there is a simple arc A_6 joining a point r on $A_4 \setminus \{q, q'\}$ with a point r' on $A_5 \setminus \{q, q'\}$ such that $A_6 \cap (A_3 \cup A_4 \cup A_5) = \{r, r'\}$. Now $A_3 \setminus \{q, q'\}$ and $A_6 \setminus \{r, r'\}$ belong to distinct arcwise connected components of $X \setminus C$ since a simple arc from A_3 to A_6 in $X \setminus C$ (having only its ends in common with A_3 and A_6) together with $A_3 \cup A_4 \cup A_5 \cup A_6$ would be an embedding of $K_{3,3}$. So, $X \setminus C$ has at least two arcwise connected components.

To see that $X \setminus C$ cannot have more than two arcwise connected components we assume (*reductio ad absurdum*) that p_1, p_2, p_3 belong to distinct arcwise connected components of $X \setminus C$. Let A_1, A_2, A_3 be pairwise disjoint segments of C. Since $X \setminus (C \setminus A_1)$ is arcwise connected it has a simple arc from p_1 to p_2. This arc contains an arc $A_{1,1}$ from p_1 to a point $q_{1,1}$ on A_1 such that $A_{1,1} \cap C = \{q_{1,1}\}$.

Similarly, there is a simple arc $A_{i,j}$ from p_i to a point $q_{i,j}$ on A_j such that $C \cap A_{i,j} = \{q_{i,j}\}$ for $i, j = 1, 2, 3$. Clearly, $A_{i,j} \cap A_{k,p} = \varnothing$ or $\{q_{i,j}\}$ when $i \neq k$. We can assume that $A_{i,j} \cap A_{i,p} = \{p_i\}$ when $j \neq p$. For $j = 1, 2, 3$, let B_j be the minimal segment of A_j that contains $q_{1,j}$, $q_{2,j}$, and $q_{3,j}$. Then the nine arcs $A_{i,j} (i, j = 1, 2, 3)$ together with $B_1 \cup B_2 \cup B_3$ form an embedding of $K_{3,3}$. This contradiction proves that $X \setminus C$ has precisely two arcwise connected components.

We shall now complete the proof by showing that (a) implies (c). We prove this by contradiction. So we assume that X contains an embedding of $K_{3,3}$. A k-cycle in $K_{3,3}$ is defined as a closed simple curve in $K_{3,3}$ consisting of k edges. We can only have $k = 4$ or 6. $K_{3,3}$ has nine 4-cycles and six 6-cycles which each separate X (since we assume (a)). Consider now a 4-cycle C. Since $K_{3,3} \setminus C$ is in an arcwise connected component of $X \setminus C$, there is another arcwise connected component which we denote by $Q(C)$. Then $Q(C)$ is also an arcwise connected component of $X \setminus K_{3,3}$. If A is a simple arc from $Q(C)$ to $K_{3,3}$ such that A has only an end q in common with $K_{3,3}$, then we say that q is a *point of attachment* of $Q(C)$ on $K_{3,3}$. All points of attachment of $Q(C)$ are on C. Moreover, if A is any segment of C, then $X \setminus (C \setminus A)$ is arcwise connected, which implies that $Q(C)$ has a point of attachment on A. In other words, the points of attachment of $Q(C)$ on $K_{3,3}$ are dense in C.

Now we may think of $K_{3,3}$ as a 6-cycle C such that the vertices of $K_{3,3}$ occur in the order $v_1, v_2, \ldots, v_6$ (where the indices are expressed modulo 6) and such that $K_{3,3}$ has the "diagonals" $v_i, v_{i+3} (i = 1, 2, 3)$. For each edge $e_i = v_i v_{i+1}$ we let C_i be the 4-cycle $v_i v_{i+1} v_{i+2} v_{i+3} v_i$, and in $Q(C_i) \cup e_i$ we select an arc A_i joining two distinct points p_i and q_i on e_i such that $A_i \cap K_{3,3} = \{p_i, q_i\}$. In C we now replace the segment B_i of e_i between p_i and q_i by A_i, $i = 1, 2, 3, 4, 5, 6$. The resulting simple curve is denoted by C'. We claim that $X \setminus C'$ is arcwise connected. $K_{3,3}$ has three 4-cycles containing two diagonals of C. For each of these there is a corresponding component of $X \setminus K_{3,3}$, and using those we conclude that all arcs B_i and all diagonals of C (except the ends) are in the same arcwise connected component R of $X \setminus C'$. In other words, $K_{3,3} \setminus C'$ is contained in R. Now suppose (*reductio ad absurdum*) that $p \in X \setminus (C' \cup R)$. Since $C' \setminus (A_1 \setminus \{p_1, q_1\})$ does not separate X, it follows that X has a simple arc D from p to A_1. Since $p \notin R$, D does not intersect $K_{3,3}$. But then $D \subseteq Q(C_1)$. In particular, $p \in Q(C_1)$. By similar reasoning we prove that $p \in Q(C_2)$. But $Q(C_1)$ and $Q(C_2)$ are disjoint. This contradiction completes the proof.

Every open connected subset of the sphere satisfies all conditions of the theorem. Since a topological space satisfying (d) of the Theorem admits no embedding of nonplanar graphs, by Kuratowski's theorem, it must be "spherelike" or "planelike." This raises the question of whether every topological space satisfying the conditions of the theorem can be embedded into the sphere.

REFERENCES

1. C. Thomassen, Kuratowski's theorem, *J. Graph Theory* 5, (1981) 225–241.
2. _____, The Jordan-Schönflies theorem and the classification of surfaces, to appear.
3. C. Thomassen and B. Toft, Nonseparating induced cycles in graphs, *J. Combinatorial Theory*, B31 (1981) 199–224.

A New Series Representation for $\zeta(3)$

JOHN A. EWELL
Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115

The Riemann zeta function ζ is defined for each complex number s having real part greater than 1 as follows:

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Actually the zeta function was first investigated by Euler about 250 years ago. Euler considered $\zeta(s)$ for real values of $s > 1$, and obtained many interesting and important results about the function. Among the more striking of these results is the now classical formula $\zeta(2) = \pi^2/6$.

In a recent issue of the MONTHLY [1] Boo Rim Choe gives yet another proof of Euler's formula, which essentially turns on integration of the power series expansion of $\sin^{-1}$. Attempting to find a closed-form expression for $\zeta(3)$ by modifying Choe's method, the author found instead a new series representation for $\zeta(3)$ that converges more rapidly than the defining series. This series representation is

$$\zeta(3) = \frac{\pi^2}{7}\left\{1 - 4\sum_{n=1}^{\infty} \frac{\zeta(2n)}{(2n+1)(2n+2)2^{2n}}\right\}. \tag{1}$$

Since $\zeta(2n) \to 1$ as $n \to \infty$, it is clear that the typical term of the series is asymptotic to $n^{-2}2^{-2n}$. Hence, the rate of convergence of this series is *much* faster than that of the defining series Σn^{-3}.

To prove (1) we consider the binomial series

$$(1 - x^2)^{-1/2} = 1 + \sum_{k=1}^{\infty} c_k x^{2k}$$

with coefficients

$$c_k = (-1)^k\binom{-1/2}{k} = 2^{-2k}\binom{2k}{k} = \frac{1 \cdot 3 \cdots (2k-1)}{2 \cdot 4 \cdots (2k)},$$

which converges for $|x| < 1$. Integration gives the series expansion of $\sin^{-1} x$:

$$\sin^{-1} x = x + \sum_{k=1}^{\infty} c_k \frac{x^{2k+1}}{2k+1}.$$

Divide the foregoing identity by x and integrate once more to obtain

$$\int_0^x t^{-1} \sin^{-1} t\, dt = x + \sum_{k=1}^{\infty} c_k \frac{x^{2k+1}}{(2k+1)^2}. \tag{2}$$

The change of variable $u = \sin^{-1} t$ transforms the integral on the left side of (2) into

$$\int_0^{\sin^{-1} x} u \cot u\, du,$$

so that when x is replaced by $\sin t$ in (2) we get

$$\int_0^t u \cot u \, du = \sum_{k=0}^{\infty} c_k \frac{\sin^{2k+1} t}{(2k+1)^2}. \tag{3}$$

But

$$u \cot u = 1 - 2 \sum_{n=1}^{\infty} \frac{\zeta(2n)}{\pi^{2n}} u^{2n},$$

which follows easily from the well-known resolution of the cotangent into partial fractions [2, p. 444]. Hence, (3) becomes

$$t - 2 \sum_{n=1}^{\infty} \frac{\zeta(2n)}{\pi^{2n}} \frac{t^{2n+1}}{2n+1} = \sum_{k=0}^{\infty} c_k \frac{\sin^{2k+1} t}{(2k+1)^2}.$$

Integrate this from 0 to $\pi/2$ and use Wallis' product [2, p. 223] in the form

$$c_k \int_0^{\pi/2} \sin^{2k+1} t \, dt = \frac{1}{2k+1}$$

to obtain

$$\frac{\pi^2}{8} - 2 \sum_{n=1}^{\infty} \frac{\zeta(2n)}{\pi^{2n}} \frac{(\pi/2)^{2n+2}}{(2n+1)(2n+2)} = \sum_{k=0}^{\infty} \frac{1}{(2k+1)^3} = \frac{7}{8}\zeta(3).$$

Simplifying this equation we thus derive (1).

The author would like to thank the referee for suggested improvements in the exposition.

REFERENCES

1. Boo Rim Choe, An elementary proof of $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$. *Amer. Math. Monthly*, 94 (1987) 662–663.
2. R. Courant, Differential and Integral Calculus, Interscience, New York, 1957.

The Dirichlet Problem for a Disk

DAVID MINDA
Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio 45221

1. Introduction. The Dirichlet problem for the open unit disk $\mathbb{D}$ can be formulated as follows. Given a continuous function $v: \partial\mathbb{D} = \mathbb{T} \to \mathbb{R}$, does there exist a continuous function $u: \overline{\mathbb{D}} \to \mathbb{R}$ such that $u|\mathbb{D}$ is harmonic and $u|\mathbb{T} = v$? To say that u is harmonic means that u is twice continuously differentiable and $\Delta u = u_{xx} + u_{yy} = 0$. The answer is affirmative and the unique solution is given by the formula

$$u(z) = \begin{cases} \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} v(e^{it}) P(z, e^{it}) \, dt & z \in \mathbb{D}, \\ v(z) & z \in \mathbb{T}. \end{cases} \tag{1}$$

Here

$$P(z, e^{it}) = \operatorname{Re}\left\{\frac{e^{it} + z}{e^{it} - z}\right\} = \frac{1 - |z|^2}{|e^{it} - z|^2}$$

denotes the Poisson kernel; $P(z, e^{it})$ is harmonic for $z \in \mathbb{D}$ and continuous for $e^{it} \in \mathbb{T}$. That u is the natural guess for the solution to the Dirichlet problem for $\mathbb{D}$ is a consequence of the Poisson integral formula.

There are two basic steps in proving that the function u defined by formula (1) actually does solve the Dirichlet problem for $\mathbb{D}$. First, it is necessary to show that u is harmonic in $\mathbb{D}$. This is relatively easy to do since it is not difficult to verify that it is permissible to differentiate under the integral sign and obtain

$$\Delta u(z) = \frac{1}{2\pi}\int_0^{2\pi} v(e^{it})\, \Delta P(z, e^{it})\, dt = 0,$$

since $P(z, e^{it})$ is harmonic as a function of $z \in \mathbb{D}$. The second step is to show that u is continuous in $\overline{\mathbb{D}}$. It is sufficient to show that u has the proper boundary values, that is,

$$\lim_{z \to e^{i\varphi}} u(z) = v(e^{i\varphi})$$

for all $e^{i\varphi} \in \mathbb{T}$. This result is called Schwarz' Theorem. The proof of this result is usually more involved because one cannot justify moving the limit operation under the integral sign in formula (1). It is worthwhile to outline one of the standard proofs of Schwarz' Theorem ([**1**, pp. 169–170], [**2**, pp. 137–139], [**3**, pp. 257–259]). First, one shows that it is sufficient to consider the special case $e^{i\varphi} = 1$ and $v(1) = 0$. Then one splits the integral into two parts (a small arc centered about 1 and the remainder of the unit circle) and notes that the first term is small since v is close to zero on the first arc, because v is continuous and vanishes at 1, while the second term is small because the Poisson kernel $P(z, e^{it})$ can be made negligible if z is close enough to 1 and e^{it} is kept a fixed distance away from 1. This is a standard "approximate identity" argument.

The purpose of this note is to present a simpler proof of Schwarz' Theorem by writing the integral in a different form so that passage to the limit under the integral sign is easily justified. This proof is known within the context of harmonic analysis and can be motivated from the viewpoint of non-Euclidean Fourier analysis [**5**, pp. 97–99]. The two main points of this article are to make this simple proof better known and to provide an elementary motivation for the proof.

2. Poisson integral formula. A motivation for the proof of Schwarz' Theorem arises naturally from one of the standard derivations of the Poisson integral formula ([**1**, pp. 166–168], [**2**, pp. 134–135], [**8**, pp. 195–196]). Let us recall this approach.

Suppose that u is continuous on $\overline{\mathbb{D}}$ and harmonic in $\mathbb{D}$. A simple extension of the mean value property for harmonic functions yields

$$u(0) = \frac{1}{2\pi}\int_0^{2\pi} u(e^{i\theta})\, d\theta. \tag{2}$$

From the point of view of geometric function theory there is a natural way to obtain an analog of formula (2) with 0 replaced by an arbitrary point $a \in \mathbb{D}$—just make use of a conformal automorphism S of $\mathbb{D}$ that sends 0 to a and replace u by $u \circ S$.

Specifically, take $S(z) = (z + a)/(1 + \bar{a}z)$ and note that S yields a homeomorphism of $\mathbb{T}$. Then $u \circ S$ is continuous in $\overline{\mathbb{D}}$ and harmonic in $\mathbb{D}$, so (2) gives

$$u(a) = (u \circ S)(0) = \frac{1}{2\pi}\int_0^{2\pi}(u \circ S)(e^{i\theta})\,d\theta = \frac{1}{2\pi}\int_0^{2\pi} u\left(\frac{e^{i\theta} + a}{1 + \bar{a}e^{i\theta}}\right) d\theta. \quad (3)$$

Now, $S^{-1}(w) = (w - a)/(1 - \bar{a}w)$ and the change of variables $e^{it} = S(e^{i\theta})$, or $e^{i\theta} = S^{-1}(e^{it})$, produces the Poisson integral formula

$$u(a) = \frac{1}{2\pi}\int_0^{2\pi} u(e^{it})\frac{1 - |a|^2}{|e^{it} - a|^2}\,dt \quad (4)$$

because $(S^{-1})'(w) = (1 - |a|^2)/(1 - \bar{a}w)^2$ yields

$$d\theta = |(S^{-1})'(e^{it})|\,dt = \frac{1 - |a|^2}{|1 - \bar{a}e^{it}|^2}\,dt = \frac{1 - |a|^2}{|e^{it} - a|^2}\,dt.$$

Here we have used $e^{it}e^{-it} = 1$ and $|e^{-it} - \bar{a}| = |e^{it} - a|$.

For our subsequent purposes the crucial observation is that

$$\frac{1}{2\pi}\int_0^{2\pi} v\left(\frac{e^{i\theta} + a}{1 + \bar{a}e^{i\theta}}\right) d\theta = \frac{1}{2\pi}\int_0^{2\pi} v(e^{it})P(a, e^{it})\,dt \quad (5)$$

for any integrable function v defined on $\mathbb{T}$.

3. Main result. We are now in a position to present a simple proof of Schwarz' Theorem.

SCHWARZ' THEOREM. *If $v : \mathbb{T} \to \mathbb{R}$ is continuous and*

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} v(e^{it})P(z, e^{it})\,dt,$$

then $\lim_{z \to e^{i\varphi}} u(z) = v(e^{i\varphi})$ for all $e^{i\varphi} \in \mathbb{T}$.

Proof. From the identity (5) we know that we can also write

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} v\left(\frac{z + e^{it}}{1 + \bar{z}e^{it}}\right) dt.$$

Now, for $e^{it} \neq -e^{i\varphi}$

$$\lim_{z \to e^{i\varphi}} \frac{z + e^{it}}{1 + \bar{z}e^{it}} = e^{i\varphi}. \quad (6)$$

Since v is continuous on the compact set $\mathbb{T}$, v is bounded on $\mathbb{T}$. The Lebesgue bounded convergence theorem implies that

$$\lim_{z \to e^{i\varphi}} u(z) = \frac{1}{2\pi}\int_0^{2\pi} \lim_{z \to e^{i\varphi}} v\left(\frac{z + e^{it}}{1 + \bar{z}e^{it}}\right) dt = \frac{1}{2\pi}\int_0^{2\pi} v(e^{i\varphi})\,dt = v(e^{i\varphi}).$$

Observe that this proof of Schwarz' Theorem works at any point of continuity of v if v is merely Lebesgue integrable and bounded on $\mathbb{T}$.

4. Comments. The one nonelementary aspect of this proof is that it makes use of Lebesgue integration. However, the bounded convergence theorem is valid for the Riemann integral and can be proved completely within the context of Riemann

integration ([4], [6], [7]). Another simple remedy can be obtained by noting that the convergence is almost uniform. The elementary inequality

$$\left| \frac{z + e^{it}}{1 + \bar{z}e^{it}} - e^{i\varphi} \right| \leqslant \frac{2|z - e^{i\varphi}|}{|e^{it} + e^{i\varphi}| - |z - e^{i\varphi}|}$$

shows that the limit (6) is uniform on compact subsets of $\mathbb{T} \setminus \{-e^{i\varphi}\}$. Since v is continuous at $e^{i\varphi}$, this implies that $v((z + e^{it})/(1 + \bar{z}e^{it}))$ converges to $v(e^{i\varphi})$ uniformly on compact subsets of $\mathbb{T} \setminus \{-e^{i\varphi}\}$. This observation together with the following elementary lemma about an almost uniformly convergent sequence of Riemann integrable functions yields a proof that avoids Lebesgue integration.

LEMMA. *Suppose that* $\{f_n\}$ *and* f *are Riemann integrable functions on* $[a, b]$. *Assume that there is a finite constant* M *such that* $|f_n(x)| \leqslant M$ *and* $|f(x)| \leqslant M$ *for all* $x \in [a, b]$ *and that for any* $\eta > 0$, *the sequence* $\{f_n\}$ *converges to* f *uniformly on* $[a + \eta, b - \eta]$. *Then*

$$\lim_{n \to \infty} \int_a^b f_n(x)\, dx = \int_a^b f(x)\, dx.$$

REFERENCES

1. L. V. Ahlfors, Complex Analysis, 3rd. ed., McGraw-Hill, New York, 1979.
2. R. B. Burckel, An Introduction to Classical Complex Analysis, vol. I, Pure and Applied Mathematics, vol. 82, Academic Press, New York, 1979.
3. J. B. Conway, Functions of One Complex Variable, 2nd. ed., Graduate Texts in Mathematics no. 11, Springer-Verlag, New York, 1978.
4. F. Cunningham, Taking limits under the integral sign, *Math. Mag.*, 40(1967) 179–186.
5. S. Helgason, Topics in Harmonic Analysis on Homogeneous Spaces, Progress in Mathematics, vol. 13, Birkhäuser, Boston, 1981.
6. H. Kestelman, Riemann integration of limit functions, *Amer. Math. Monthly*, 77(1970) 182–187.
7. J. W. Lewin, A truly elementary approach to the bounded convergence theorem, *Amer. Math. Monthly*, 93(1986) 395–397.
8. R. Nevanlinna and V. Paatero, Introduction to Complex Analysis, Addison-Wesley, Reading, Mass., 1969.

Diagonalization over Commutative Rings

R. BRUCE RICHTER
Deptartment of Mathematics, Carleton University, Ottawa, Ontario, Canada K1S 5B6

WILLIAM P. WARDLAW
Department of Mathematics, U.S. Naval Academy, Annapolis, Maryland 21402

1. Introduction. We are interested in diagonalizing matrices over the integers: given an integral matrix A, find an integrally invertible matrix P such that $P^{-1}AP$ is diagonal or show that no such matrix exists.

In this note, we solve this problem. To obtain our characterization (given here as Theorem 5), we prove a variety of elementary generalizations to commutative rings of theorems that are well known over fields.

There is an extensive literature on the topic of linear algebra over commutative rings. The texts [2, 3, 4] are all valuable references. However, we found they lack

many of the results we required and had expected to find in such texts. In particular, we could not find our Theorems 1 and 5 in the literature.

It is our purpose to present this material in an elementary manner which is accessible to the reader having only a basic knowledge of linear algebra and an introductory knowledge of rings.

2. Basics. There are many similarities between the development of the portion of linear algebra presented here and the usual development over fields. However, the absence of inverses and the existence of divisors of zero create some problems.

Example 1. Let $A = \begin{bmatrix} 1 & 0 \\ 4 & 2 \end{bmatrix}$. If $P = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}$, then P^{-1} is integral and $P^{-1}AP$ is the diagonal matrix diag(1, 2). Hence A is diagonalizable over the integers $\mathbf{Z}$. Let $\mathbf{Z}_n$ be the integers modulo n. Then A is evidently diagonalizable over $\mathbf{Z}_n$ for each integer $n > 1$. However, if $Q = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$, then, over $\mathbf{Z}_6$, $Q^{-1}AQ = \text{diag}(4, 5)$, so the diagonal form of A is not unique up to ordering of the diagonal elements. Note that $\mathbf{Z}_6$ is the direct sum of the fields $\mathbf{Z}_2$ and $\mathbf{Z}_3$. Over the former, diag(4, 5) = diag(2, 1), while over the latter, diag(4, 5) = diag(1, 2). Thus, in a sense, the diagonal elements are the same, but reordered.

Henceforth, we shall let R be a commutative ring with unity 1 and M an R-module. (See [**1**] for definitions; we assume M is unital, i.e., $1v = v$ for every $v \in M$.) *Independent sets*, *spanning sets* and *bases* in M are defined as usual for sets of vectors over a field. Warning: Not every module has a basis.

(We have chosen to use the words independent and spanning rather than free and generating—as in [**3**]—to emphasize the parallel we are intending to the theory over fields. The definitions given for free and generating in [**3**] allow greater flexibility in certain developments with which we shall not be concerned.)

A module is *free* if it has a basis. The module $R^{n\times p}$ of all $n \times p$ matrices with entries in R is an example of a free module. We shall be dealing chiefly with $R^n = R^{n\times 1}$ and its submodules.

We begin with some relevant facts about R-modules. Suppose $V = (v_1, \ldots, v_k)$ and $W = (w_1, \ldots, w_n)$ are sequences of elements of the module M and $A = (\alpha_{ij})$ is an $n \times k$ matrix over R. Then we write $V = WA$ if $v_j = \sum_{i=1}^{n} \alpha_{ij} w_i$, for $j = 1, \ldots, k$. If M is a submodule of R^d, then W can be interpreted as a $d \times n$ matrix and WA as a matrix product. (We speak here of sequences of vectors rather than sets of vectors, since the ordering is important.)

We have not been able to find the following result in [**2**], [**3**], and [**4**]. On the other hand, [**3**], for example, does show that any two bases of a free module have the same cardinality.

THEOREM 1. *Let V and W be the sequences above and suppose V is independent and W is spanning. Then $k \leq n$. Moreover, $k = n$ implies W is a basis (but V need not be).*

Proof. Since W spans, $V = WA$ for some $A \in R^{n\times k}$. If $k > n$, then [**2**, Thm. 51] implies there is a non-zero $x \in R^k$ such that $Ax = 0$, contradicting the independence of V.

Suppose $k = n$. We must show W is independent. Now $x \neq 0$ implies $Ax \neq 0$, since V is independent. Thus, [**2**, Cor. to Thm. 51] implies $\det(A)$ is not a divisor of zero; so [**2**, Thm. 50; or **6**, Thm. 3] implies A has an inverse B over the quotient ring

Q of R. (The denominators in Q are the non-divisors of zero in R, so $\det(A)$ is a unit in Q.)

If $Wx = 0$ for some $x \in R^n$, then $0 = W(AB)x = (WA)Bx = Vu$, where $u = Bx \in Q^n$. Multiplying u by δ to clear the denominators yields $y = \delta u = \delta Bx \in R^n$. Then $Vy = \delta Vu = 0$, which implies $y = 0$. Now $\delta x = \delta(AB)x = \delta Au = Ay = 0$, so $x = 0$, since δ is not a divisor of zero. Thus, W is independent. ■

Several corollaries to Theorem 1 or its proof can now be easily established, via arguments well known over fields.

COROLLARY 1.1. *Let P and Q be matrices over R, each of whose columns form a basis of the same subspace of R^n. Then there is an R-invertible matrix S such that $P = QS$.*

COROLLARY 1.2. *If P is an $n \times n$ matrix over R, then the columns of P form a basis for R^n if and only if P is invertible over R (if and only if $\det(P)$ is a unit in R).*

Let $A \in R^{n \times n}$. An element λ of R is an *eigenvalue* of A if there is a nonzero vector $v \in R^n$ such that $Av = \lambda v$. The vector v is an *eigenvector*.

COROLLARY 1.3. *Let a be an $n \times n$ matrix over R. Then A is diagonalizable over R if and only if there is a basis of R^n consisting of eigenvectors of A.*

3. Eigenvalues and eigenvectors. Corollary 1.3 is central to our discussion. Thus, we must find a basis of eigenvectors of A. Now λ is an eigenvalue of A if and only if the matrix $C = C_A(\lambda) = \lambda I - A$ is singular (i.e., $Cv = 0$ for some $v \neq 0$) if and only if $f_A(\lambda) = \det[C_A(\lambda)]$ is a divisor of 0.

If v is a member of a basis of eigenvectors, then $\{v\}$ is independent, i.e., $\alpha v = 0$ implies $\alpha = 0$. If $Av = \lambda v$, then, for $C = C_A(\lambda)$, $Cv = 0$, so $\operatorname{adj}(C)Cv = \det(C)v = 0$ implies $\det(C) = 0$. Thus, $f_A(\lambda) = 0$. It follows that we need only consider eigenvectors corresponding to zeroes of the characteristic polynomial of A. In summary, we have

LEMMA 2. *Let A be an $n \times n$ matrix over R with characteristic polynomial $f_A(t)$. Then $\lambda \in R$ is an eigenvalue of A if and only if $f_A(\lambda)$ is a divisor of zero. Moreover, the eigenvalue λ has an associated independent eigenvector only if $f_A(\lambda) = 0$.*

Another comment is in order. In [3] it is observed that an independent vector v need not be part of a basis of a free module M. (For example, 2 is independent in $\mathbf{Z}$, but is in no basis of $\mathbf{Z}$.) A vector v is *unimodular* if it is independent and $M = \langle v \rangle \oplus N$ for some submodule N of M, where $\langle v \rangle = \{rv | r \in R\}$. If v is in some basis of M, then v is unimodular. Thus, the "moreover" statement of Lemma 2 might have read, "The eigenvalue λ has an associated unimodular eigenvector only if $f_A(\lambda) = 0$."

We observe that if R has any nonzero nilpotent elements and $n > 1$, then the converse to the moreover statement of Lemma 2 fails. For example, if $A = \operatorname{diag}(3, 3)$ over $\mathbf{Z}_4$, then $f_A(1) = 0$, but no eigenvector corresponding to 1 is independent.

For the eigenvalue λ of A, let $E_\lambda = \{v | Av = \lambda v\}$ be the associated eigenmodule. Let $\Omega(A) = \{\lambda \in R | f_A(\lambda) = 0\}$. To find a basis of eigenvectors of A, we need only consider elements of the E_λ for $\lambda \in \Omega(A)$.

Example 2. For the matrix A from Example 1, over $R = \mathbf{Z}_6$, every element of R is an eigenvalue! However, $f_A(\lambda) = 0$ only for $\lambda = 1, 2, 4$ and 5, so $\Omega(A) = \{1, 2, 4, 5\}$. For each $\lambda \in \Omega(A)$, let $S_\lambda = \{v_\lambda\}$ be a maximal independent subset of E_λ. Of the six pairs of these four sets, only two pairs have unions which are independent, namely $S_1 \cup S_2$ and $S_4 \cup S_5$. These give us the two diagonalizations seen in Example 1.

Example 2 shows the naiveté of the definition of eigenvalue. One should, perhaps, define λ to be an eigenvalue if there is an independent (or unimodular) vector v such that $Av = \lambda v$. In fact, in [3] an eigenvector is required to be in some basis of the module!

The following is the commutative ring version of a well-known result for fields; the proof is the usual induction and is omitted.

LEMMA 3. *Suppose* $\Omega \subseteq \Omega(A)$ *has the property that* $\lambda, \tau \in \Omega$ *and* $\lambda \neq \tau$ *imply* $\lambda - \tau$ *is not a divisor of zero in* R. *Then the sum* $\Sigma_{\lambda \in \Omega} E_\lambda$ *is direct. Hence,* $S_\lambda \subseteq E_\lambda$ *independent implies* $\bigcup_{\lambda \in \Omega} S_\lambda$ *is independent.*

The matrix $A = \operatorname{diag}(2, 4)$ over $\mathbf{Z}_6$ shows the converse of Lemma 3 is false.

Lemma 3 explains the outcome of Example 2; more generally it provides a mechanism for constructing "large" independent sets of eigenvectors.

4. Diagonalizing. Since the converse of Lemma 3 can fail, the general problem of building a basis of eigenvectors seems to be quite difficult. However, when R is an integral domain, Lemma 3 shows that the sum of the eigenmodules is direct. The procedure for diagonalization would seem to be: (1) find a basis of each eigenmodule; and (2) if the union of these bases (an independent set) is a basis of R^n, then A is diagonalizable. Obviously, two problems may arise: (1) some eigenmodule may have no basis; and (2) the union of the particular bases we choose might not be a basis of R^n. (We note that if R is a principal ideal domain, then every module has a basis, so problem (1) does not arise [4, Thm. 230].)

Example 3. Consider the matrix $A = \begin{bmatrix} \sqrt{-5} & 2 \\ 0 & 1 \end{bmatrix}$ over the ring $R = \mathbf{Z}[\sqrt{-5}]$. The eigenvalues of A are $\sqrt{-5}$ and 1, but the eigenmodule E_1 has no basis. Thus, this is an example of problem (1).

Example 4. Consider the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ over $\mathbf{Z}$. The reader should have little difficulty showing that each eigenmodule has a basis, but no union of their bases is a basis of Z^2. This, then, is an example of problem (2).

Perhaps somewhat surprisingly, these are the only problems that can arise in diagonalizing over integral domains.

THEOREM 5. *Let* R *be an integral domain and let* $A \in R^{n \times n}$. *Then* A *is diagonalizable over* R *if and only if each eigenmodule of* A *has a basis over* R *and, for every choice of these bases their union is a basis of* R^n.

Of course, as the special case of interest, we have the following.

COROLLARY 5.1. *Let* $A \in \mathbf{Z}^{n \times n}$. *Then* A *is diagonalizable over* Z *if and only if, for any choice of bases of the (integral) eigenspaces of* A, *the union of these bases is a basis of* Z^n.

Proof of Theorem 5. If there is any basis of eigenvectors, Corollary 1.3 shows A is diagonalizable over R.

Now suppose A is diagonalizable over R. Let $Q^{-1}AQ = D$, where Q is invertible over R and D is a diagonal matrix. The eigenvalues of A and their multiplicities are determined in the quotient field K of R. These are the entries of D. Without loss of generality, we may assume that equal eigenvalues of A occur in consecutive columns of D in the order $\lambda_1, \ldots, \lambda_k$. Write $Q = [Q_1 \cdots Q_k]$, where the columns of Q_j form a basis over K of the eigenmodule associated with λ_j.

We show now that Q_j is a basis over R of the eigenmodule associated with λ_j. If $v \in R^n$ and $Av = \lambda_j v$, then $v = Q(Q^{-1}v)$ is the unique expression of v as an R-linear combination of the columns of Q (the columns of Q form a basis of R^n). This is also the unique expression of v as a K-linear combination of the columns of Q. But v is in the eigenspace (over K) associated with λ_j, so v is a K-linear combination of the columns of Q_j. Hence the columns of Q_j form an R-basis of the R-eigenmodule associated with λ_j. Thus, each eigenmodule of A (over R) must have a basis.

Suppose that, for each j, P_j is a matrix whose columns form an R-basis of the eigenmodule associated with λ_j and let $P = [P_1 \cdots P_k]$. By Corollary 1.1, for each j there is an invertible matrix S_j such that $P_j = Q_j S_j$. Let S be the block diagonal matrix $\operatorname{diag}(S_1, \ldots, S_k)$. Evidently, S is invertible over R. Since $P = QS$ is the product of invertible matrices, P is also invertible over R. It follows from Corollary 1.2 that the columns of P form a basis of R^n. Hence, the union of the arbitrarily selected bases P_j is a basis of R^n. ■

To actually find bases of the eigenmodules E_λ, we suppose R is a Euclidean domain. The bases always exist [**4**, Thm. 230] and can be found by the algorithm summarized below.

Let $C = C_A(\lambda) = \lambda I - A$. Column operations making repeated use of the division algorithm can be used to put the gcd of the first row of C in the (1, 1)-position and zero every other entry of the first row. Proceeding inductively and keeping track of the elementary matrices producing the column operations, we can find a matrix U which is invertible over R such that CU is column reduced to lower triangular form. CU is called the *Hermite Normal Form* (HNF) of C. All the zero columns of CU are to the right and it is not difficult to show that the columns of U in the same positions as the zero columns of CU form a basis of the eigenmodule E_λ of A. (Doing this in polynomial time over the integers is described in [**5**]. We are grateful to W. J. Cook for having brought the HNF to our attention.)

Corollary 5.1 and the procedure just described realize our goal of determining which integral matrices are diagonalizable over the integers and diagonalizing those that are.

REFERENCES

1. I. N. Herstein, Topics in Algebra, John Wiley and Sons, New York, 1975.
2. N. H. McCoy, Rings and Ideals, Carus Monograph No. 8, Math. Assoc. of America, 1948.
3. B. R. McDonald, Linear Algebra Over Commutative Rings, Pure and Applied Mathematics #87, Marcel Dekker, New York, 1984.
4. L. Redei, Algebra, Vol. 1, Pergamon Press, Oxford, 1967.
5. A. Schrijver, Theory of Linear and Integer Programming, Wiley, Chichester, 1986.
6. W. P. Wardlaw, A transfer device for matrix theorems, *Math. Mag.*, 59 (1986) 30–33.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

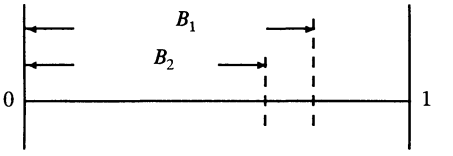
Random Triangles

JOE WHITTAKER
Department of Mathematics, University of Lancaster, Lancaster, LA1 4YL, United Kingdom

A question similar to the one that follows occurred on the final examination of a second-year course in probability at the University of Birmingham. One wonders if the setter got it right.

> *A chimpanzee named Euclid is teaching himself the geometry of the triangle by notching thin, straight jungle twigs with his finger nail at two independent random points, breaking the twig into three at these points, and then noting if it is possible to construct a triangle from the fragments. Nearby Pythagoras, a colleague of Euclid's, is learning geometry in a similar fashion; his procedure is the same except that after making the first notch he breaks the twig into two and then etches the second notch on the longer of the two resulting portions. After breaking this, he too tries to construct a triangle from the three fragments.*
>
> *Archimedes, a rather learned chimpanzee, strolls through the jungle clearing and observes his two companions. After a moment's reflection he comments, "Pythagoras, you are* 4/3 *times as likely as Euclid to succeed in constructing a triangle."*
>
> *Geometrically or otherwise evaluate the two probabilities and say whether Archimedes's calculation is correct.*

Analysis. Consider an idealized twig of unit length and random break-points B_1, B_2 measured from one end of the twig.



The sample space for this experiment can be parameterized by the set of all pairs of break-points, (b_1, b_2), falling in the unit square (see Figure 1). A success, where a triangle is possible, occurs if and only if the sum of any two sides is greater than the third. When $b_2 < b_1$, the lengths of the three sides are b_2, $b_1 - b_2$ and $1 - b_1$, and a success implies: (i) $b_2 + b_1 - b_2 > 1 - b_1$, so that $b_1 > 1/2$; (ii)$b_2 + 1 - b_1 > b_1 - b_2$ so that $b_1 - b_2 < 1/2$; and (iii) $b_1 - b_2 + 1 - b_1 > b_2$, so that $b_2 < 1/2$. These three linear inequalities determine one of the two triangular regions enclosed by the dashed lines in Figure 1(a). The other is determined by consideration of the

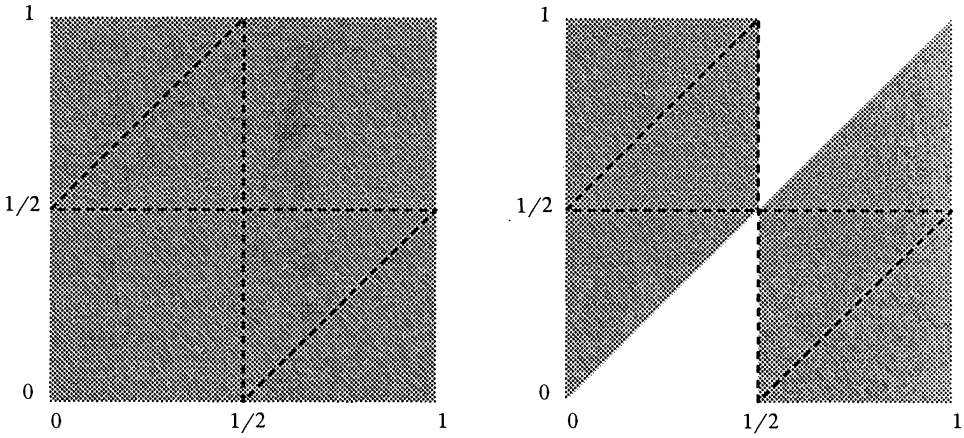


FIG. 1. The sample space.

case $b_2 > b_1$. Exterior to the dotted triangles, the square regions are failures as the two notches fall in the same half of the twig, and the triangular regions are failures as the two notches are too far apart. For Euclid, all points in the square are possible, and by symmetry equal areas are equally probable, so that his probability of success is $1/4$. Now Pythagoras's sample space is a subset of Euclid's modified to exclude the region corresponding to making the second break in the shorter fragment. So from Figure 1(b), Pythagoras's success probability is $1/3$ and Archimedes is right.

But there is a twist to the tale: under Pythagoras's sampling scheme, equal areas are *not* equally probable. To see this, suppose they are and calculate the areas in the vertical strips of width δ located on the horizontal axis at two points $0 < \alpha < \beta < 1/2$. From Figure 1(b) the probability $P(B_1$ lies in $\alpha \pm \delta/2)$ is greater than $P(B_1$ lies in $\beta \pm \delta/2)$, but this violates the uniform distribution of B_1 along the twig.

To ensure a uniform distribution we have to assign a larger probability to the shorter interval: so given $B_1 = b_1$ $(> 1/2)$ say, then B_2 is uniformly distributed on the interval $(0, b_1)$ and not the interval $(0, 1)$. Hence

$$P\left(B_2 \text{ lies in } \alpha \pm \frac{\delta}{2} | B_1 = b_1\right) = \frac{\delta}{(b_1 - 0)}$$

and the density function is illustrated in Figure 2. Consequently Pythagoras's success probability is

$$2\int_{b_1=1/2}^{1}\left\{\int_{b_2=b_1-1/2}^{1/2}\frac{1}{b_1}db_2\right\}db_1$$

which simplifies to

$$2\int_{1/2}^{1}\frac{1}{b_1}(1 - b_1)db_1 = 2[\log b_1 - b_1]_{1/2}^{1} = 2\log 2 - 1 = 0.38\ldots$$

and not $0.33\cdots$, Archimedes's analysis was naive.

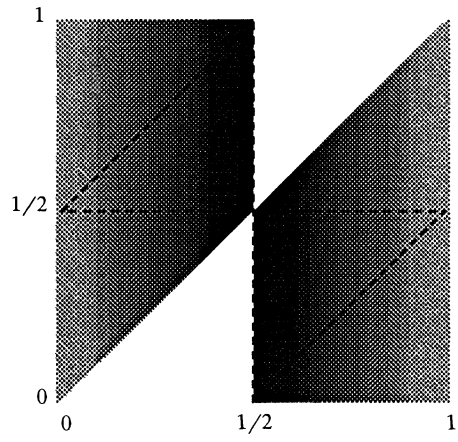


FIG. 2. Unequal probabilities. Pythagoras density function.

A New Extension of the Derivative

DANIEL KOPEL
10 Polo Road, Great Neck, NY 11023

MICHAEL SCHRAMM
Department of Mathematics, LeMoyne College, Syracuse, NY 13214

How, in an elementary setting, can we go beyond calculating one example after another to learn about differentiation? Can we see what differentiation *is* rather than just how it *works*? What are the vital properties that an operator must have to be called a "derivative?" One way to consider these questions is to construct extensions of the derivative. In checking whether a proposed extension is what we want it to be, we learn more than we could by just doing examples. In this note we examine one such extension. The motivation for our definition is very simple, and might form the basis of other, similar experiments.

The relationship of the graph of a function to its tangent line is described in familiar words: "Look closer and closer at the point of tangency, and the curve and its tangent line look more and more alike." Most methods of extending the derivative retain the difference quotient as the measure of how much a curve "looks like" its tangent line, and proceed by manipulating the way the limit of the difference quotients is computed (see e.g., [**1, 2**]).

We take a different approach: Reinterpret the phrase "The curve looks like the line." There are many ways in which a set of points can be said to "look like" a line. The one given by the ordinary definition of the derivative may not be the most natural. On the other hand, most of us can recognize the similarity of two sets of points, even as we appreciate the difficulties in defining it.

One way in which a set of points may be said to look like a line is in the sense of least squares. Usually used to find the line that best fits a finite set of points, the

method may be adapted for our purposes in this way: Let $m(h)$ and $b(h)$ be the numbers which minimize the expression

$$\int_x^{x+h} [f(t) - m(h)t - b(h)]^2 dt$$

for a given value of h (we are concerned only with $m(h)$, the slope of the approximating line). Setting partial derivatives to 0, solving the resulting system of equations, and integrating where possible, we find

$$m(h) = \frac{12}{h^3}\left\{\int_x^{x+h} tf(t)\,dt - \left[x + \frac{h}{2}\right]\int_x^{x+h} f(t)\,dt\right\}.$$

We say f is *Least Squares Differentiable* if $\lim_{h\to 0} m(h)$ exists, and call this quantity the *Least Squares Derivative* of f at x, denoted $\mathrm{LSD}_x f$.

We now ask some questions concerning this operator. Some of these questions are specific to the Least Squares Derivative, while others can be asked, with appropriate modifications, of any proposed extension of the derivative (in what follows, words such as "differentiable" and "derivative" without qualification mean "differentiable in the ordinary sense" and "the ordinary derivative," etc.).

Is it linear? Any operator that is to be called a derivative certainly must be linear. The integrals which define $m(h)$ act linearly on the function f, and the limit acts linearly on $m(h)$, so LSD_x is linear.

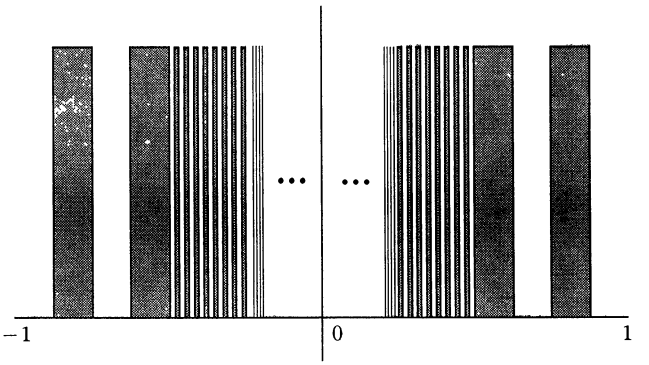
Is it an extension of the derivative? We want our operator to be an *extension* of the derivative, that is, any function that is differentiable should be least squares differentiable, and $\mathrm{LSD}_x f$ should be the same as $f(x)$ in that case. We will omit most of the details, but recall that the function f is differentiable at the point a with derivative L if $f(x) - f(a) - L(x - a) = \varepsilon(x - a)$, where $\varepsilon \to 0$ as $x \to a$. We may substitute $f(a) + L(x - a) + \varepsilon(x - a)$ for $f(x)$ in the integrals defining $m(h)$ and obtain the desired result.

Is it a proper extension of the derivative? Are there any functions that are least squares differentiable but not differentiable? The affirmative answer to this is quite easy, and reveals our motivation for considering this extension in the first place. Change the value of a differentiable function f at the point at which the derivative is to be computed. This will render f nondifferentiable, but the integrals used to evaluate $m(h)$ are not effected by such a change. The function will still be least squares differentiable. If we use Lebesgue integrals, we may change the function on a set of measure 0 without losing least squares differentiability. It would seem that the least squares derivative is an improvement over the derivative in this sense. This leads us to our next question.

Are there any least squares differentiable functions that do not arise in the way just described? If there were none, our "new" derivative would be of little interest. The example we give here is more complicated than it need be to answer this question, but will be of use later: Let C_1 be the characteristic function of the set

$$\left\{x: \frac{1}{2^{i+1}} + \frac{2j}{8^{i+1}} \le |x| \le \frac{1}{2^{i+1}} + \frac{2j+1}{8^{i+1}} i = 0,1,\ldots,\ j = 0,1,\ldots,2^{2i+1} - 1\right\}.$$

Here we have shaded, in part, the region under the graph of C_1:



It may be shown that $\mathrm{LSD}_0 C_1 = 0$, and that C_1 is not equal almost everywhere to a differentiable function.

This example works because the integrals do not distinguish well between comb-like functions with very thin teeth and functions that are slowly increasing. They ignore some of the essential character of such a function, which is apparent in our next results.

Does it satisfy other familiar differentiation rules? While we may not wish to make this requirement, the extent to which a new operator mimics (or fails to mimic) the behavior of a familiar one is certainly of interest. Here we must adjust our expectations. We can't expect the product rule, $\mathrm{LSD}_x(fg) = f(x)\mathrm{LSD}_x g + g(x)\mathrm{LSD}_x f$, to hold in a setting where we may change the values $f(x)$ and $g(x)$ in any way we wish. We might try to avoid this difficulty by substituting, for instance,

$$\lim_{h\to 0} \frac{1}{h} \int_x^{x+h} f(x)\,dx$$

for $f(x)$ (the relationship between $m(h)$ and $b(h)$ makes this seem appropriate, and the limit is equal to $f(x)$ if f is continuous). But this does not help. Let $C_2(x) = xC_1(x)$, then $\mathrm{LSD}_0 C_2 = \frac{1}{2}$.

$$\lim_{h\to 0} \frac{1}{h} \int_0^h C_1(x)\,dx = \frac{1}{2} \quad \text{and} \quad \lim_{h\to 0} \frac{1}{h} \int_0^h C_2(x)\,dx = 0.$$

Now $C_2(x) = C_1(x)C_2(x)$, so even our modified product rule would yield $\frac{1}{2} = \mathrm{LSD}_0 C_2 = \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 = \frac{1}{4}$. Similarly, LSD_x fails to satisfy the chain rule: Since $C_2(x) = C_2(C_2(x))$, we would have

$$\tfrac{1}{2} = \mathrm{LSD}_0 C_2 = \mathrm{LSD}_0(C_2 \circ C_2) = \mathrm{LSD}_0 C_2 \cdot \mathrm{LSD}_0 C_2 = \tfrac{1}{4}.$$

Further questions suggest themselves: In view of the last result, should we consider this attempt a failure? If we *extend* the derivative (thereby widening the class of "differentiable" functions) we should expect to lose some of its properties. How important are these lost properties?

We know that "Differentiable ⇒ Least Squares Differentiable." Is there a simple property that can be inserted to make "Least Squares Differentiable and ???? ⇒ Differentiable" true? ("Continuous" will not work. The function C_2 is continuous at 0, but is not differentiable there.)

How does this derivative behave with regard to other properties and uses of the derivative? For instance, is there a meaningful version of Taylor's Theorem for this derivative?

Most important, what does this preliminary exploration teach us about the derivative? While suggesting some directions for experimentation, these results contain a warning that familiar properties of the derivative might be more strongly attached to the "limit of difference quotients" than we suspect. This brings us full circle to one of our very first questions: What are the vital properties that an operator must have to be called a "derivative"? This is probably where the discussion should begin!

REFERENCES

1. C. L. Belna, M. J. Evans, and P. D. Humke, Symmetric and strong differentiation, this MONTHLY, 86 (1979) 121–123.
2. A. M. Bruckner and J. L. Leonard, Derivatives, this MONTHLY, 73 (1966) 24–56.

Remainder Estimates in Taylor's Theorem

G. B. FOLLAND
Department of Mathematics, University of Washington, Seattle, WA 98195

Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is n times differentiable at a, that is, f is of class $C^{(n-1)}$ in a neighborhood of a and $f^{(n)}(a)$ exists. Then we can form the nth degree Taylor polynomial and remainder of f at a:

$$P_n(h, a) = \sum_0^n \frac{f^{(j)}(a)}{j!} h^j, \qquad R_n(h, a) = f(a + h) - P_n(h, a).$$

The versions of Taylor's theorem in almost all calculus books, both elementary and advanced, give some formula for R_n that involves the $(n + 1)$st derivative of f. Actually, it is quite unnecessary to invoke $f^{(n+1)}$ to obtain good control of R_n. This point is perhaps too fine for standard first-year courses, but it should be of interest in honors or advanced calculus courses.

In the first place, assuming only that f is n times differentiable at a, $R_n(h, a)$ is $o(h^n)$ as $h \to 0$, that is,

$$\lim_{h \to 0} \frac{R_n(h, a)}{h^n} = 0. \tag{1}$$

Indeed, after an $(n - 1)$-fold application of l'Hôpital's rule, (1) reduces to

$$\lim_{h \to 0} \frac{f^{(n-1)}(a + h) - f^{(n-1)}(a)}{h} - f^{(n)}(a) = 0,$$

which is true by definition of $f^{(n)}(a)$. (An analogous result is valid for functions of several variables, but it cannot be obtained quite so cheaply: see [**1**].)

Under the slightly stronger hypothesis that f is of class $C^{(n)}$ on an interval containing a, we can obtain better bounds for R_n as follows. We start with the

usual integral formula for R_{n-1},

$$R_{n-1}(h, a) = \frac{h^n}{(n-1)!}\int_0^1 (1-t)^{n-1} f^{(n)}(a+th)\,dt,$$

which is easily proved by integration by parts, and give it an extra twist:

$$\begin{aligned} R_n(h, a) &= R_{n-1}(h, a) - \frac{f^{(n)}(a)}{n!}h^n \\ &= \frac{h^n}{(n-1)!}\int_0^1 (1-t)^{n-1}\left[f^{(n)}(a+th) - f^{(n)}(a)\right]dt. \end{aligned} \tag{2}$$

Now any estimate one may have for the modulus of continuity of $f^{(n)}$ at a can be translated into an estimate for $R_n(h, a)$. For example:

(I) If $f^{(n)}$ is continuous on an interval I, then $f^{(n)}(a+th) - f^{(n)}(a) \to 0$ as $h \to 0$ uniformly for t in $[0, 1]$ and a in any compact subset of I; hence (1) holds uniformly for a in compact subsets of I.

(II) If $f^{(n)}$ is Hölder continuous of exponent λ, $0 < \lambda \leqslant 1$, then $R_n(h, a)$ is dominated by $|h|^{n+\lambda}$. More precisely, if

$$|f^{(n)}(a+x) - f^{(n)}(a)| \leqslant C|x|^\lambda \tag{3}$$

for all x on the line segment from 0 to h, then

$$|R_n(h, a)| \leqslant \frac{C|h|^{n+\lambda}}{(n-1)!}\int_0^1 (1-t)^{n-1}t^\lambda dt = \frac{C|h|^{n+\lambda}}{(1+\lambda)(2+\lambda)\dots(n+\lambda)}. \tag{4}$$

(To prove the last equality, integrate by parts.) In particular, this yields the standard estimate for R_n under the hypothesis that f is of class $C^{(n+1)}$ and $|f^{(n+1)}| \leqslant C$, for then (3) holds with $\lambda = 1$ by the Mean Value Theorem.

The same arguments work equally well for functions of several variables, say f: $\mathbb{R}^p \to \mathbb{R}$. (We denote points in $\mathbb{R}^p$ by $\mathbf{x} = (x_1, \dots, x_p)$.) Suppose f is of class $C^{(n)}$. Since $(d/dt)f(\mathbf{a} + t\mathbf{h}) = \mathbf{h} \cdot \nabla f(\mathbf{a} + t\mathbf{h})$, the nth Taylor polynomial of f is

$$P_n(\mathbf{h}, \mathbf{a}) = \sum_{j=0}^{n} \frac{(\mathbf{h} \cdot \nabla)^j f(\mathbf{a})}{j!}, \tag{5}$$

and the integral formula for the remainder analogous to (2) is

$$R_n(\mathbf{h}, \mathbf{a}) = \frac{1}{(n-1)!}\int_0^1 (1-t)^{n-1}\left[(\mathbf{h} \cdot \nabla)^n f(\mathbf{a} + t\mathbf{h}) - (\mathbf{h} \cdot \nabla)^n f(\mathbf{a})\right]dt. \tag{6}$$

However, it is preferable to restate these formulas so that the partial derivatives of f appear more explicitly. To this end, we employ multi-index notation: $\alpha = (\alpha_1, \dots, \alpha_p)$ denotes a p-tuple of nonnegative integers, and

$$|\alpha| = \alpha_1 + \cdots + \alpha_p, \qquad \alpha! = \alpha_1! \cdots \alpha_p!,$$

$$\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_p^{\alpha_p}}, \qquad \mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_p^{\alpha_p}.$$

With this notation, we have the "multinomial theorem,"

$$\sum_{|\alpha|=k} \frac{k!}{\alpha!}\mathbf{x}^{\alpha} = (x_1 + \cdots + x_p)^k, \tag{7}$$

which can be proved from the binomial theorem by induction on p. If we use this formula to expand the operator $(\mathbf{h} \cdot \nabla)^k$, (5) and (6) become

$$P_n(\mathbf{h}, \mathbf{a}) = \sum_{|\alpha|\leqslant n} \frac{\partial^{\alpha} f(\mathbf{a})}{\alpha!}\mathbf{h}^{\alpha},$$

$$R_n(\mathbf{h}, \mathbf{a}) = n \sum_{|\alpha|=n} \frac{\mathbf{h}^{\alpha}}{\alpha!} \int_0^1 (1-t)^{n-1}[\partial^{\alpha} f(\mathbf{a} + t\mathbf{h}) - \partial^{\alpha} f(\mathbf{a})]\, dt. \tag{8}$$

Now information about the derivatives $\partial^{\alpha} f$, $|\alpha| = n$, can be used to estimate R_n. The analogue of (4) works out most neatly if we use the norm on $\mathbb{R}^p$ defined by

$$|\mathbf{x}| = |x_1| + \cdots + |x_p|.$$

Indeed, if we have

$$|\partial^{\alpha} f(\mathbf{a} + \mathbf{x}) - \partial^{\alpha}(\mathbf{a})| \leqslant C|\mathbf{x}|^{\lambda}$$

for $|\alpha| = n$ and $\mathbf{x}$ on the line segment from $\mathbf{0}$ to $\mathbf{h}$, (7) and (8) yield

$$|R_n(\mathbf{h}, \mathbf{a})| \leqslant nC|\mathbf{h}|^{\lambda} \sum_{|\alpha|=n} \frac{|\mathbf{h}^{\alpha}|}{\alpha!} \int_0^1 (1-t)^{n-1} t^{\lambda} dt = \frac{C|\mathbf{h}|^{n+\lambda}}{(1+\lambda)\cdots(n+\lambda)}.$$

I am grateful to the referee for replacing my original argument with a simpler one.

REFERENCE

1. J. Facenda Aquirre, A note on Taylor's theorem, this MONTHLY, 96 (1989) 244–47.

Symmetry in Logos and Hubcaps

JOSEPH A. GALLIAN
Department of Mathematics and Statistics, University of Minnesota, Duluth, MN 55812

In common things that round us lie
Some random truths he can impart.
William Wordsworth, *The Prelude*

In [7] and [8] Hungarian needlework was used to illustrate the frieze groups and the crystallographic groups. Others have used African, Chinese, Indian, and Islamic art to illustrate the same families. But what about finite groups of symmetries? What are rich sources for illustrating cyclic and dihedral groups through symmetry? My favorite ones are logos and hubcaps. These ubiquitous artifacts are preferable to exotic artworks and crafts for educational purposes because students encounter logos and hubcaps daily. Leafing through the yellow pages or strolling through a parking lot becomes a lesson in group theory!

A font for logos is the fascinating book *American Trademark Designs* [3]. Recall Leonardo's Theorem: The only finite plane symmetry groups are Z_n and D_n, the cyclic group of order n and the dihedral group of order $2n$ [5, p. 375]. Perusing [3] and the yellow pages for Madison, Wisconsin, I found corporate logos with the cyclic symmetry groups Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8, Z_{16}, and Z_{20} and the dihedral symmetry groups D_1, D_2, D_3, D_4, D_5, D_6, D_8, D_{10}, D_{12}, D_{14}, D_{16}, D_{20}, and D_{24}. Many of these are illustrated in FIGURES 1 and 2. Note that although Z_2 and D_1 are isomorphic as groups they are geometrically distinct.

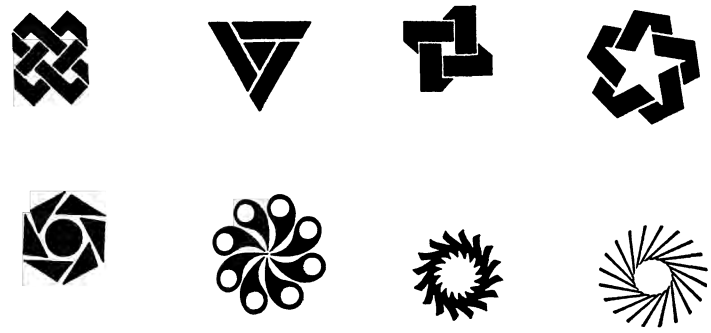


FIG. 1. Logos with symmetry groups Z_2, Z_3, Z_4, Z_5, Z_6, Z_8, Z_{16}, and Z_{20}.

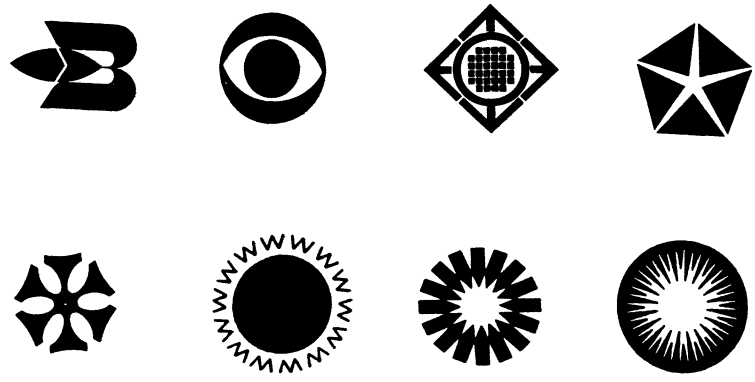


FIG. 2. Logos with symmetry groups D_1, D_2, D_4, D_5, D_6, D_{14}, D_{16}, and D_{20}.

Automobile hubcaps and wheels offer an even more varied array of finite symmetry groups provided one makes allowances for imperfections such as the notch for the valve stem, the automobile name or logo, and the area where the wheel lugs are. Leonardo's Theorem applies here as well since the symmetry group of the orthogonal projection of a hubcap onto a plane (e.g., the photographic image) is isomorphic to the three-dimensional symmetry group of the hubcap itself. (A reflection across a plane in three dimensions corresponds to a reflection across a line

in the orthogonal projection.) In one brief foray to a few parking lots I netted hubcaps (or wheels) with Z_n symmetry groups for n = 2, 5, 9, 10, 12, 15, 17, 18, 20, 22, 24, and D_n symmetry groups for n = 2, 3, 4, 5, 6, 7, 8, 10, 12, 13, 15, 16, 18, 20, 24, 30, 32, 34, 36, and 72. Some of these are shown in FIGURES 3 and 4. (Some photos were shot at a slight angle.) Subsequent searching turned up hubcaps with symmetry groups Z_n for n = 3, 4, 6, 7, 8, 14, 16, 31, and 60 and D_n for n = 28, 35, 40, 76, 80, and 100 (yes, 100).

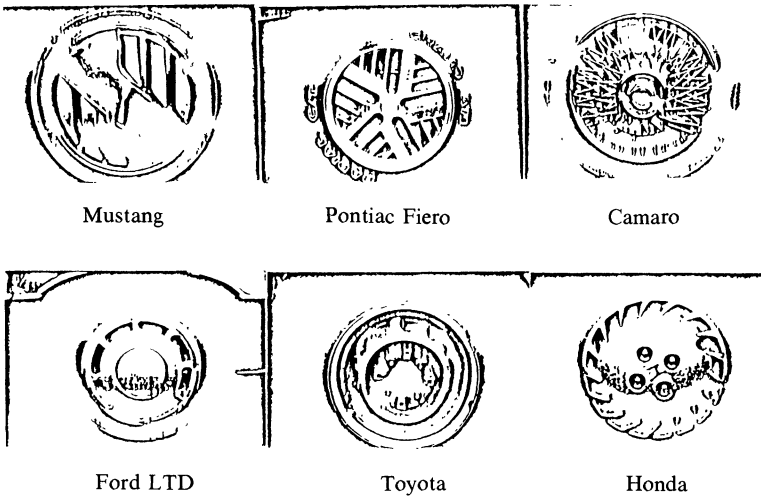


FIG. 3. Hubcaps with symmetry groups Z_2, Z_5, Z_6, Z_8, Z_{12}, and Z_{17}.

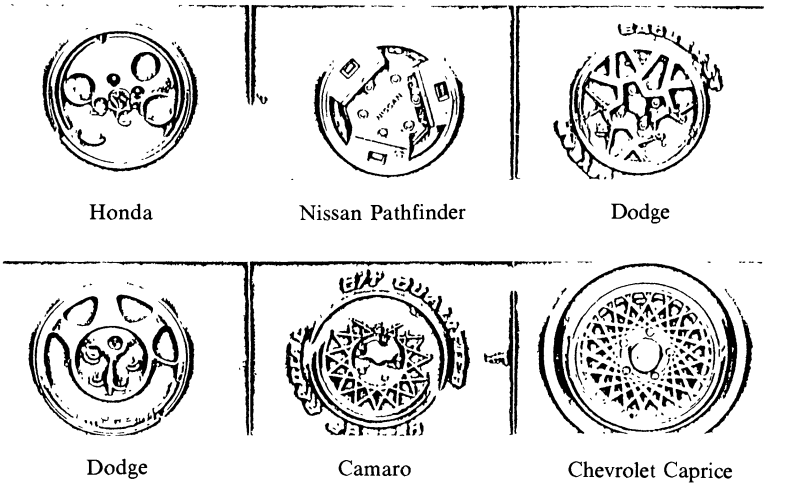


FIG. 4. Hubcaps with symmetry groups D_2, D_3, D_4, D_7, D_{15}, and D_{20}.

REFERENCES

1. Leroy H. Appleton, American Indian Design & Decoration, Dover, New York, 1971.
2. W. and G. Audsley, Designs and Patterns from Historic Ornament, Dover, New York, 1968.
3. Barbara B. Capitman, American Trademark Designs, Dover, New York, 1976.
4. P. Cowen, Rose Windows, Chronicle Books, San Francisco, 1979.
5. J. A. Gallian, Contemporary Abstract Algebra, 2nd ed., D. C. Heath, Lexington, 1990.
6. I. Hargittai and M. Hargittai, Symmetry through the Eyes of a Chemist, VCH, New York, 1986.
7. I. Hargittai and G. Lengyel, The seven one-dimensional space-group symmetries illustrated by Hungarian folk needlework, *J. of Chemical Education*, 61 (1984) 1033–1034.
8. I. Hargittai and G. Lengyel, The seventeen two-dimensional space-group symmetries in Hungarian needlework, *J. Chemical Education*, 62 (1985) 35–36.
9. A. V. Shubnikov and V. A. Koptsik, Symmetry in Science and Art, Plenum Press, New York, 1974.
10. Peter S. Stevens, Handbook of Regular Patterns, MIT Press, Cambridge, 1980.
11. Dorothy K. Washburn and D. W. Crowe, Symmetries of Culture: Handbook of Plane Pattern Analysis, University of Washington Press, Seattle, 1988.
12. H. Weyl, Symmetry, Princeton University Press, Princeton, 1952.

in closed form (in terms of a finite number of trigonometric and hyperbolic functions).

E 3376. *Proposed by Robert J. Blodgett, Morningside, MD.*

Prove that

$$\sum_{i=0}^{N}\sum_{j=0}^{N}\binom{i+j}{j}^2\binom{4N-2i-2j}{2N-2j} = (2N+1)\binom{2N}{N}^2$$

for any positive integer N.

E 3377. *Proposed by Farhoud Pouryoussefi (student), Sharif University of Technology, Tehran, Iran.*

Suppose we consider the polygon with vertices $e^{2\pi i\theta_1}, e^{2\pi i\theta_2}, \ldots, e^{2\pi i\theta_n}$ in the complex plane, where

$$0 \leqslant \theta_1 < \theta_2 < \cdots < \theta_n < 2\pi.$$

Suppose α and β are given positive numbers with $\alpha + \beta = 1$. Define $\theta_i^{(0)} = \theta_i$ for $i = 1, 2, \ldots, n$ and

$$\theta_i^{(k)} = \alpha\theta_i^{(k-1)} + \beta\theta_{i+1}^{(k-1)}$$

for $k = 1, 2, \ldots$, where subscripts are taken modulo n and the angles $\theta_i^{(k)}$ are taken modulo 2π. Prove that

$$\lim_{k\to\infty}\left(\theta_{i+1}^{(k)} - \theta_i^{(k)}\right) = 2\pi/n$$

for $i = 1, 2, \ldots, n$.

E 3378. *Proposed by Miklòs Bòna, Budapest, Hungary.*

Suppose the points of Z^2 (the set of points in the plane with integer coordinates) are colored with a finite number of colors. For every $n \geqslant 3$ prove that there exists a convex n-gon with vertices and centroid in Z^2 such that all $n + 1$ points have the same color.

SOLUTIONS OF ELEMENTARY PROBLEMS

Sums of Five Composite Squares

E 3262 [1988, 351]. *Proposed by R. G. E. Pinch, Emmanuel College, Cambridge, and C. Boyd, University of Edinburgh.*

It is known that every natural number n can be expressed as the sum of four squares of integers, and that three squares suffice unless n is of the form $4^a(8b + 7)$. Show that every sufficiently large natural number can be expressed as the sum of at most five squares of composite numbers (i.e., squares of positive integers divisible either by the square of a prime or by the product of two distinct primes).

Solution I by L. E. Mattics, University of South Alabama, Mobile, AL. Let A be any natural number not less than $4(2 \cdot 5^4 + 1)^2 + 144$. We may choose a number x from the set $\{6, 10, 12\}$ such that

$$A - x^2 \equiv \pm 4 (\mathrm{mod}\, 5).$$

Since the equation

$$A - x^2 \equiv 4z^2 (\mathrm{mod}\, 5)$$

is solvable for z, there is an integer w such that

$$A - x^2 \equiv 4w^2 (\mathrm{mod}\, 5^4), \qquad 2 \le w \le 5^4 + 1.$$

(Cf. Niven and Zuckerman, *An Introduction to the Theory of Numbers*, New York, 1960, 1966, 1972, 1980, §2.6.) By taking y as either w or $w + 5^4$ we may choose an integer y satisfying the three conditions

$$A - x^2 \equiv 4y^2 (\mathrm{mod}\, 5^4), \quad A - x^2 - 4y^2 \not\equiv 0, 7 (\mathrm{mod}\, 8), \quad 2 \le y \le 2 \cdot 5^4 + 1.$$

Since $(A - x^2 - 4y^2)/5^4$ is a non-negative integer congruent to $A - x^2 - 4y^2$ modulo 8, there are integers a, b, c such that

$$(A - x^2 - 4y^2)/5^4 = a^2 + b^2 + c^2.$$

(Cf. Dickson, *Modern Elementary Theory of Numbers*, Chicago, 1939, §53.) Hence

$$A = x^2 + (2y)^2 + (25a)^2 + (25b)^2 + (25c)^2,$$

where x and $2y$ are composite positive integers and $25a$, $25b$, and $25c$ are either zero or composite.

Solution II, prepared by the editors on the basis of a solution by the late Emil Grosswald, Temple University, Philadelphia, PA. We prove that every sufficiently large positive integer is a sum of exactly five squares of composite numbers.

Let $r_s(n)$ be the number of solutions of $x_1^2 + x_2^2 + \cdots + x_s^2 = n$ in integers and let $r_s^*(n)$ be the number of solutions of the same equation in which $|x_i|$ is a composite positive integer for each i.

It is known (Theorem 386 of Hardy and Wright) that

$$r_4(n) = 8 \sum_{d \mid n,\, 4 \nmid d} d,$$

so that $r_4(n) \ge 8n$ if n is not divisible by 4. Thus

$$\begin{aligned} r_5(n) &= \sum_{-\sqrt{n} \le j \le \sqrt{n}} r_4(n - j^2) \\ &\ge \sum_{\substack{-\sqrt{n/2} \le j \le \sqrt{n/2} \\ j \equiv n+1 (\mathrm{mod}\, 2)}} r_4(n - j^2) \\ &\ge \sum_{\substack{-\sqrt{n/2} \le j \le \sqrt{n/2} \\ j \equiv n+1 (\mathrm{mod}\, 2)}} 4n \\ &\ge 4n\left(\sqrt{n/2} - 1\right) \ge 2n^{3/2}, \end{aligned}$$

provided $n \ge 25$.

Now $r_5(n) - r_5^*(n)$ is the number of solutions of $x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2$ in which at least one of the summands is 0, ± 1, or $\pm p$ for some prime number p. Thus

$$r_5(n) - r_5^*(n) \leq 5r_4(n) + 10r_4(n-1) + 10 \sum_{p \leq \sqrt{n}} r_4(n - p^2).$$

Now

$$r_4(n) \leq 8 \sum_{d \mid n} d = 8n \sum_{d \mid n} \frac{1}{d} = O(n \log \log n)$$

by Theorem 323 of Hardy and Wright. Hence

$$r_5(n) - r_5^*(n) = O(n \log \log n) + O\big(\pi(\sqrt{n})n \log \log n\big),$$

where $\pi(x)$ is the number of primes not exceeding x. By Chebychev's estimate $\pi(x) = O(x/\log x)$ we get

$$r_5(n) - r_5^*(n) = O\big(n^{3/2}(\log \log n)/\log n\big).$$

Combining the assertions of the preceding two paragraphs, we find that $r_5^*(n) > n^{3/2}$ for all sufficiently large n.

Editorial comment. The proposer remarked that there are exactly 256 positive integers which cannot be expressed as a sum of at most five squares of composite numbers and that the largest of these is 1167.

Several mathematicians have pointed out to the editors that actually every sufficiently large positive integer can be expressed as a sum of at most four squares of composite numbers. This can be established in several ways, none of them elementary. If $n \equiv 0 (\mathrm{mod}\, 16)$ and n is expressed as a sum of four squares of non-negative integers, all of the integers involved are divisible by 4 and hence the non-zero ones among them are composite. If n is a sufficiently large positive integer not divisible by 16, then by the results on pp. 453–454 of [**1**] it follows that n is represented by the quadratic form

$$(69x)^2 + (95y)^2 + (119z)^2 + (143w)^2.$$

Thus every sufficiently large n is expressible as a sum of at most four squares of composite numbers.

REFERENCE

1. H. D. Kloosterman, On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$, *Acta. Math.*, 49(1926) 407–464.

Solved also by the proposers.

More Socks in the Laundry

E 3265 [1988, 456]. *Proposed by D. M. Friedlen, Georgia Institute of Technology, Atlanta, GA.*

Suppose n distinct pairs of socks are put into a laundry, where it is assumed that each of the $2n$ socks has precisely one mate. At the completion of the laundering operation, socks are drawn out one at a time. Suppose the first pair is realized on draw number T_n.

(a) Find an explicit formula for $E(T_n)$, the expected value of T_n, and determine

$$\lim_{n\to\infty} n^{-1/2}E(T_n).$$

(b) For positive x, prove that

$$\lim_{n\to\infty} P\{T_n \le xn^{1/2}\}$$

exists and find its value.

Solution by Doug Pryor, University of California, Santa Barbara, CA. For (a), we show that $E(T_n) = 2^{2n}/\binom{2n}{n}$, from which an application of Stirling's formula yields $\lim n^{-1/2}E(T_n) = \sqrt{\pi}$. Note that $E(T_n) = \sum_{k=2}^{n+1} kP(T_n = k) = \sum_{k=0}^{n} P(T_n > k)$. However, $P(T_n > k) = \binom{n}{k}2^k/\binom{2n}{k}$, because $T_n > k$ means that, of the $\binom{2n}{k}$ ways to select the first k socks, we have chosen exactly one from each of some k pairs. After canceling and rearranging terms, this yields $E(T_n) = (n!)^2(2n)!^{-1}\sum_{k=0}^{n}\binom{2n-k}{n}2^k$. These summands are precisely 2^{2n} times the probabilities in the well-known Banach match box problem (see, e.g., W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. I, 3rd ed., Wiley, 1968, p. 166–167), and thus they sum to 2^{2n}.

For (b), we show that $\lim P(T_n \le x\sqrt{n}) = 1 - e^{-x^2/4}$. Let $k = \lfloor x\sqrt{n} \rfloor$. Using the formula obtained in (a), we have $P(T_n \le x\sqrt{n}) = 1 - P(T_n > k) = 1 - \binom{n}{k}2^k/\binom{2n}{k}$. By applying Stirling's formula and rearranging factors approximately, we can obtain an asymptotic expression for $P(T_n > k)$, namely,

$$P(T_n > k) = \frac{2^k n!}{(n-k)!}\frac{(2n-k)!}{(2n)!}$$

$$\sim \frac{2^k n^n}{(n-k)^{n-k}}\frac{(2n-k)^{2n-k}}{(2n)^{2n}} = \left\{\frac{1-k/n}{1-k/(2n)}\right\}^k \left\{\frac{(1-k/(2n))^2}{1-k/n}\right\}^n.$$

We claim that the two factors in the last expression approach $e^{-x^2/2}$ and $e^{x^2/4}$, respectively, which will complete the proof. In fact the logarithm of the first factor is

$$k\{\log(1-k/n) - \log(1-k/(2n))\} = -k^2/(2n) + O(k^3/n^2),$$

which approaches $-x^2/2$ as n goes to infinity, while the logarithm of the second factor is

$$n\log\left\{\frac{1-k/n+k^2/(4n^2)}{1-k/n}\right\} = \frac{k^2}{4(n-k)} + O\left(\frac{k^4}{n(n-k)^2}\right),$$

which approaches $x^2/4$.

Editorial comment. Many readers noted the connection with the Banach match box problem; others gave ad hoc proofs of the summation. Richard Groenveld

remarks that the asymptotic distribution $P(T \leq x) = 1 - e^{-x^2/4}$ found in (b) is known as the Weibull distribution.

A related problem about socks in the laundry appeared as E 3148 [1986, 400; 1988, 357].

Kent Kromarek remarked on the empirical aspects of this problem.

Solved also by A. Adler, G. Bach (West Germany), C. Belise and J. Sarkar, D. A. Callan, S. Dharmadhikari, S. Ethier and P. Griffin, R. A. Groenveld, E. Hertz, L. Holst (Sweden), L. E. Mattics, W. A. Newcomb, G. Parker, N. Passell, D. S. Rubin, A. Shimron (Israel), T. R. Shore, P. Tzermias (Greece), and the proposer.

The Limit of an Approximation

E 3276 [1988, 654]. *Proposed by D. M. Bloom, Brooklyn College, CUNY.*

D. J. Newman (*A Problem Seminar*, Springer-Verlag, 1982, Problem 60) shows that if (x_n) is the sequence defined by $x_1 = 1$ and $x_{n+1} = x_n + 1/x_n$ for $n \geq 1$, then

$$\sqrt{2n} < x_n < \sqrt{2n + \tfrac{1}{2}\ln n}$$

for $n \geq 3$. Put

$$G(n) = 2n + \tfrac{1}{2}\ln n - x_n^2$$

for $n \in Z^+$. Show that maximum and minimum values of $G(n)$ exist, find these extrema, and show that $\lim_{n\to\infty} G(n)$ exists.

Solution by Kee-Wai Lau, Hong Kong. The maximum and minimum values of $G(n)$ are, respectively, $G(1) = 1$ and $G(13) \approx 0.2671$. Computation shows that the value of $G(n)$ decreases when n goes from 1 to 13, but that $G(13) < G(14)$. We shall prove that $G(n)$ increases but is bounded (by 0.305) when n goes from 14 to ∞, so that $\lim_{n\to\infty} G(n)$ exists.

We shall use mathematical induction to prove the more specific result that, if $n \geq 14$, then

$$0 < G(13) < G(14) < \cdots < G(n) \leq G(13) + \sum_{k=13}^{n-1} (\ln k)/(8k^2). \qquad (1)$$

So suppose $n \geq 14$ and that we know that (1) holds. From this inductive assumption we shall deduce that

$$0 < G(n+1) - G(n) < (\ln n)/(8n^2) \qquad (2)$$

and consequently

$$0 < G(13) < \cdots < G(n) < G(n+1) < G(13) + \sum_{k=13}^{n} (\ln k)/(8k^2). \qquad (3)$$

We first note that the equality $x_{n+1}^2 = x_n^2 + 2 + 1/x_n^2$ implies

$$G(n+1) - G(n) = \frac{1}{2}\ln\left(1 + \frac{1}{n}\right) - \frac{1}{2n + \tfrac{1}{2}\ln n - G(n)}. \qquad (4)$$

Since

$$G(13) + \sum_{k=13}^{\infty} \frac{\ln k}{8k^2} < G(13) + \int_{12}^{\infty} \frac{\ln x}{8x^2}\,dx = G(13) + \frac{\ln 12 + 1}{96} < 0.305,$$

our inductive assumption (1) gives $\frac{1}{2}\ln n - G(n) > \frac{1}{2}\ln 14 - 0.305 > 1$; thus by (4) we have

$$G(n+1) - G(n) > \frac{1}{2}\ln\left(1 + \frac{1}{n}\right) - \frac{1}{2n+1} > 0.$$

On the other hand, our inductive assumption (1) trivially gives $G(n) > 0$ and so by (4)

$$G(n+1) - G(n) < \frac{1}{2}\ln\left(1 + \frac{1}{n}\right) - \frac{1}{2n + \frac{1}{2}\ln n} < \frac{1}{2n} - \frac{1}{2n + \frac{1}{2}\ln n} < \frac{\ln n}{8n^2}.$$

Thus we have deduced (2) and (3). Thus completes the induction and establishes the results of the problem.

Editorial comment. According to Jean Anglesio, the numerical value of $\lim G(n)$ is 0.2768576....

Solved also by J. Anglesio (France), D. Borwein (Canada), D. Callan, Chico Problem Group, Z. Franco, S. M. Gagola, Jr., S. Gaignoux (France), R. Martin (West Germany), J.-M. Monier (France), M. Vowe (Switzerland), and the proposer.

A Property of Singular Matrices

E 3277 [1988, 654]. *Proposed by Tim Sauer, George Mason University, Fairfax, VA.*

Suppose A and B are n by n matrices and suppose there exists a nonzero vector x such that $Ax = 0$ and a vector y such that $Ay = Bx$. If A_i is the matrix obtained from A by replacing its ith column by the ith column of B, show that

$$\sum_{i=1}^{n} \det A_i = 0.$$

Solution I by Thomas Jager, Calvin College, Grand Rapids, MI. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be the columns of A and B, and let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be the entries of x and y. By symmetry, we may assume $x_1 \neq 0$. Since the determinant is a linear function of the first column, we may write

$$x_1 \sum_{i=1}^{n} \det A_i = \det(x_1 b_1, a_2, \ldots, a_n) + \sum_{i=2}^{n} \det(x_1 a_1, a_2, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n).$$

Since $\Sigma x_i a_i = 0$ and the determinant is an alternating multilinear function of all its

columns, for $i \geq 2$ we have

$$\begin{aligned}\det(x_1a_1, a_2, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n) &= \sum_{j=2}^{n} \det(-x_ja_j, a_2, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n) \\ &= \det(-x_ia_i, a_2, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n) \\ &= \det(x_ib_i, a_2, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n)\end{aligned}$$

Translating $Ay = Bx$ as $\Sigma x_ib_i = \Sigma y_ia_i$, we have

$$\begin{aligned}x_1 \Sigma \det A_i &= \Sigma \det(x_ib_i, a_2, \ldots, a_n) = \det\Big(\Sigma y_ia_i, a_2, \ldots, a_n\Big) \\ &= \det(y_1a_1, a_2, \ldots, a_n) = y_1 \det A = 0,\end{aligned}$$

since the existence of x implies that A is singular. Dividing by $x_1 \neq 0$ completes the proof.

Solution II by A. A. Jagers, Universiteit Twente, Enschede, The Netherlands. The condition $Ax = 0$ implies $\det A = 0$ and rank $A < n$. If rank $A < n - 1$, then each A_i is singular and $\Sigma \det A_i = 0$, so we may assume rank $A = n - 1$, and the nullspace of A is spanned by x.

Let $\tilde{A}$ be the adjoint matrix of A; i.e., the i, jth entry of $\tilde{A}$ equals $(-1)^{i+j}$ times the determinant of the matrix obtained by omitting row j and column i of A. If we expand each $\det A_i$ using cofactors along column i, we obtain $\Sigma \det A_i = \operatorname{trace}(\tilde{A}B)$. Now $\tilde{A}A = (\det A) \cdot I = 0 \cdot I$ implies that $A(\tilde{A}B) = (\det A) \cdot B = 0$. Hence each column of $\tilde{A}B$ is in the nullspace of A and must be a multiple of x. Since we also have $\tilde{A}Bx = \tilde{A}Ay = (\det A)y = 0$, we conclude $(\tilde{A}B)^2 = 0$. This implies that every eigenvalue of $\tilde{A}B$ is zero, so $\operatorname{trace}(\tilde{A}B) = 0$, which completes the proof.

Solution III by Albert Nijenhuis, Seattle. Let R be the commutative ring $\mathbb{R}[t]/(t^2)$; its elements have the form $a + tb$, where $a, b \in \mathbb{R}$, $t^2 = 0$, and $a + tb = 0$ if and only if $a = b = 0$. The advantage of computations in R is that any square matrices A, B with real-valued entries satisfy $\det(A + tB) = \det A + t\Sigma \det A_i$. Let C be the adjoint of $A + tB$ over R, defined as in the preceding solution but using computations in R; since R is commutative, $C(A + tB) = \det(A + tB) \cdot I$ still holds. The conditions $Ax = 0$ and $Ay = Bx$ imply that, over R, $(A + tB)(x - ty) = 0$. Also, $Ax = 0$ implies $\det A = 0$. Putting all these facts together, we have the following computation over R:

$$\begin{aligned}0 = C(A + tB)(x - ty) &= \det(A + tB)(x - ty) = (\det A + t\Sigma \det A_i)(x - ty) \\ &= (\det A)x + t\big[-(\det A)y + (\Sigma \det A_i)x\big] = t(\Sigma \det A_i)x.\end{aligned}$$

Because x is non-zero, this requires $\Sigma \det A_i = 0$, as desired.

Solution IV by Yan-Loi Wong, Colby College, Waterville, ME. As in Solution I, let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ denote the columns of A and B. We may assume by symmetry and scaling that $x_1 = 1$. Then $Ay = Bx$ and $Ax = 0$ imply for any real t that

$$\begin{aligned}\det(A + tB) &= \det((A + tB)x, a_2 + tb_2, \ldots, a_n + tb_n) \\ &= t \det(Ay, a_2 + tb_2, \ldots, a_n + tb_n).\end{aligned}$$

This yields

$$\frac{d}{dt}\bigg|_{t=0} \det(A + tB) = \det(Ay, a_2, \ldots, a_n) = \det(\Sigma y_i a_i, a_2, \ldots, a_n) = y_1 \det A = 0.$$

However, we also have $\frac{d}{dt}\Big|_{t=0} \det(A + tB) = \Sigma \det A_i$ for any square A and B, so $\Sigma \det A_i = 0$.

Solved also by S. J. Bernau, J. T. Bruening, D. Callan, D. G. Cantor, J. Ferrer (Spain), S. M. Gagola, Jr., L. N. Howard, W. Janous (Austria), L. D. Kadison, O. P. Lossers (The Netherlands), J.-M. Monier (France), P. Montgomery, A. Müller (Switzerland), J. H. Nieto (Venezuela), A. Pedersen (Denmark), P. P. Ray, P. A. Staley, J. H. Steelman, N. S. Thornber, L. J. Wallen, S. Zakeri (Iran), Western Maryland College Problems Group, and the proposer. Three incorrect solutions were received.

Disjoint Arithmetic Progressions

E 3278 [1988, 654]. *Proposed by John Beebee, Bakersfield, CA.*

Let $a \bmod m$ denote the arithmetic progression $a + km$: $k \in Z$. Suppose $b \bmod n$ is disjoint from $a \bmod m$. What is the smallest positive integer r such that there is an arithmetic progression $c \bmod r$ which is disjoint from both $a \bmod m$ and $b \bmod n$?

Solution I by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. The key observation is that $a \bmod m$ and $b \bmod n$ are disjoint if and only if $a \not\equiv b \pmod d$, where $d = gcd(m, n)$. Thus r must not be relatively prime to either m or n. Let p, q, and s be the smallest primes dividing m, n, and d, respectively. The necessary condition just stated for r implies $r \geq \min\{s, pq\}$. To complete the determination of r, we consider several cases.

If $s > 2$, then the answer is $r = \min\{s, pq\}$. Because $a \bmod m$ and $b \bmod n$ are contained in progressions $\bmod s$ and there are at least three of these, there is a progression $c \bmod s$ disjoint from $a \bmod m$ and $b \bmod n$. On the other hand, if $pq < s$, then of the pq different progressions $\bmod pq$, at most q intersect $a \bmod m$ and at most p intersect $b \bmod n$. Since $p + q < pq$, there is a progression $c \bmod pq$ disjoint from both.

If $s = 2$ and $a \equiv b \pmod 2$, then the answer is $r = 2$. Here both $a \bmod m$ and $b \bmod n$ are contained in the same progression $\bmod 2$, and the other is disjoint from them.

If $s = 2$ and $a \not\equiv b \pmod 2$, then the answer is $r = \min\{s', 2t\}$, where s' is the smallest odd prime dividing d and t is the smallest prime dividing either $m/2$ or $n/2$. The fact that r must not be relatively prime to m or n gives the lower bound, and an analysis like that for $s > 2$ above gives the existence of such a sequence.

Solution II by Shippensburg University Mathematical Problem Solving Group, Shippensburg, PA. The answer is the smallest element of the set

$$\{[s, t]: s, t > 1, s|m, t|n, [s, t] > 3 - (a - b, 2)\},$$

where $[x, y]$ denotes the least common multiple of x and y and (x, y) denotes the greatest common divisor of x and y.

To see this, we first observe that $a \bmod m \cap b \bmod n = \varnothing$ if and only if $a - b \neq ln - km$ for all integers k, l, which holds if and only if (m, n) does not divide $a - b$. Let us say that r is *good* if there exists an integer c such that $c \bmod r$ is disjoint from $a \bmod m$ and $b \bmod n$. For a good r, let $d_1 = (r, m)$ and $d_2 = (r, n)$. Then d_1 does not divide $a - c$ and d_2 does not divide $b - c$. Since $([d_1, d_2], m) = d_1$ and $([d_1, d_2], n) = d_2$, it follows that $[d_1, d_2]$ is good. We also have $[d_1, d_2] \leq r$, so we need consider only $S = \{[s, t]: s, t > 1, s|m, t|n\}$.

Every element of S is good except possibly the element 2 when $([s, t], m) = ([s, t], n) = 2$. In this situation $[s, t] = 2$ and 2 is good if and only if $a - b$ is even.

Solved also by A. Averill, D. Callan, N. Felsinger, J. Ferrer (Spain), S. Gaignoux (France), D. E. Manes, H. M. Marston, A. Nijenhuis, A. Pedersen (Denmark), J. L. Selfridge, R. W. Sheets, and the proposer. Partially solved by J.-M. Monier (France). Two incorrect solutions were received.

An Extremal Problem for Triangle-Free Graphs

E 3284 [1988, 762]. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest, Hungary.*

Determine the minimum number of edges in a triangle-free graph on $2n$ vertices whose complement contains no complete graph on n vertices.

Solution by Richard Stong, Harvard University, Cambridge, MA. For $n \leq 3$ there are no such graphs. For $n = 4$ the minimum number of edges in such a graph is 10, and for $n \geq 5$ it is $n + 5$. For $n = 4$ the minimum is realized by an 8-cycle $v_1, \ldots, v_8$ with two consecutive diagonals v_1v_5 and v_2v_6; call this graph G_4. For $n \geq 5$ the minimum is realized by the graph G_n consisting of $n - 5$ isolated edges and two disjoint 5-cycles.

An *independent* set is a set of vertices no two of which are adjacent; forbidding an n-clique from the complement of G is equivalent to considering graphs G with no independent set of size n. For $n \geq 5$, an independent set in G_n uses at most one vertex from each isolated edge and at most two vertices from each 5-cycle, for a total of at most $n - 1$ vertices. For $n = 4$, four vertices in G_4 are independent only if they alternate on the 8-cycle, but the diagonals prevent those two sets in G_4 from being independent. For $n = 1, 2$, it is clear that there is no such graph, and for $n = 3$ it is well known that any graph on 6 vertices has a triangle or an independent set of size 3. It remains to show that the examples above cannot be improved.

For simplicity, call a graph n-good if it is a triangle-free graph G on $2n$ vertices with no independent set of size n, and let $f(n)$ denote the minimum number of edges in an n-good graph. Let G be an n-good graph with $f(n)$ edges. From the desired properties, we may assume that G has no loops or multiple edges.

If G has a vertex v of degree at most 1, then there is an $(n - 1)$-good graph, and $f(n) \geq f(n - 1) + 1$. To see this, let w be the neighbor of v if it has one, and otherwise let w be an arbitrary non-isolated vertex. Any independent set in $G - \{v, w\}$ can be augmented in G by adding v, and deleting vertices cannot introduce triangles. Thus, $G - \{v, w\}$ is $(n - 1)$-good, and the loss of edge(s) incident to w shows $f(n) \geq f(n - 1) + 1$.

If $f(n) < n + 5$ and $n \geq 5$, then the average of the vertex degrees in G is less than 2, and G has a vertex of degree at most 1. Thus the proof is completed by

proving that $f(4) = 10$. Because there is no 3-good graph, we may assume every vertex of G has degree at least 2. If G_4 is not optimal, then the optimal G has 8 or 9 edges and degree sequence 2, 2, 2, 2, 2, 2, 2, 2 or 3, 3, 2, 2, 2, 2, 2, 2. There are two such triangle-free graphs with 8 edges and 6 with 9 edges, but each of these has an independent set of size 4.

Editorial comment. If the condition "triangle-free" is removed from the statement, then the problem is solved by the well-known theorem of P. Turán (see, for example, J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, (North-Holland, 1976), p. 109–111). Turán's Theorem states that the maximum number of edges in an N-vertex graph having no complete graph on r vertices is attained *uniquely* by the complete multipartite N-vertex graph with $r - 1$ parts in which any two part-sizes differ by at most 1. If $N = 2n$ and $r = n$, the complement of this graph is the disjoint union of two triangles and $n - 3$ edges, which has $n + 3$ edges.

Solved also by S. F. Barger, M. Chlebik (Czechoslovakia), R. B. Eggleton (Brunei), N. Hartsfield and J. Werth, and P. Tracy. Three other submissions gave the correct answer but no proof.

A Determinant Identity with Matrix Entries

E 3286 [1988, 763]. *Proposed by William C. Waterhouse, Pennsylvania State University, State College, PA.*

Suppose A, B, C, D are n by n matrices and C', D' denote the transposes of C, D. In a comment on problem 6057 [1987, 1020], M. J. Pelling remarked that the condition

$$CD' + DC' = 0 \tag{*}$$

implies

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(AD' + BC') \tag{**}$$

if D is nonsingular, but gave an example with singular D satisfying (*) in which one side of (**) is 1 and the other side is -1. Prove that in any case (*) implies

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix}^2 = \det(AD' + BC')^2.$$

Solution I by Dennis R. Estes, University of Southern California, Los Angeles, CA. Let I_n denote the n by n identity matrix, and let $P = \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}$. Then

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix}^2 = \det\begin{pmatrix} A & B \\ C & D \end{pmatrix}\det\left(P\begin{pmatrix} A & B \\ C & D \end{pmatrix}'P\right) = \det\left(\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} D' & B' \\ C' & A' \end{pmatrix}\right)$$

$$= \det\begin{pmatrix} AD' + BC' & AB' + BA' \\ 0 & CB' + DA' \end{pmatrix}$$

$$= \det(AD' + BC')\det(AD' + BC')' = \det(AD' + BC')^2.$$

Solution II and generalization by David Callan, University of Bridgeport, Bridgeport, CT. We will show that (*) implies

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = (-1)^r \det(AD' + BC'), \qquad (***)$$

where $r = \operatorname{rank}(C)$. Since both hypothesis and conclusion are unchanged when A, B, C, D are replaced respectively by $Q^{-1}AQ, Q^{-1}B(Q')^{-1}, PCQ, PD(Q')^{-1}$ for suitable nonsingular matrices P, Q, we may assume that $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ has the matching block form

$$\begin{pmatrix} A_{11} & A_{12} & B_{11} & B_{12} \\ A_{21} & A_{22} & B_{21} & B_{22} \\ I & 0 & D_{11} & D_{12} \\ 0 & 0 & D_{21} & D_{22} \end{pmatrix},$$

where I is the r by r identity matrix. The hypothesis implies $D_{11}' = -D_{11}$ and $D_{21} = 0$. We have the following simple rules for manipulating the determinant of a blocked matrix. If a (matrix) multiple of one row of blocks is added to another, then the determinant is unchanged. If two adjacent rows of blocks (involving r_1, r_2 rows respectively of the full matrix) are interchanged, then the determinant is multiplied by $(-1)^{r_1 r_2}$. Similarly for column manipulations. Using the rules, it is easy to manipulate both sides of (***) into the same form: $(-1)^{r \cdot n} \det D_{22} \det\begin{pmatrix} A_{12} & B_{11} - A_{11}D_{11} \\ A_{22} & B_{21} - A_{21}D_{11} \end{pmatrix}$.

Editorial comment. A collaborating editor required a full handwritten page for each of the "manipulations" mentioned in the last sentence of Solution II.

In [1] Marvin Marcus proved a result involving the same determinants as (* * *) but with $(-1)^r$ replaced by $(-1)^{n-s}$, where s is the rank of D. In fact, if (*) holds and if $\det(AD' + BC') \neq 0$, then $n - s \equiv r \pmod 2$.

1. Marvin Marcus, Two determinant condensation formulas, *Linear and Multilinear Algebra*, 22 (1987) 95–102.

Solved also by R. C. Alperin, D. Callan, S. Gaignoux (France), Y. Ikeda, A. A. Jagers (The Netherlands), B. Kirchheim (Czechoslovakia), O. Krafft (West Germany), J. R. Kuttler, O. P. Lossers (The Netherlands), T. L. Markham, J.-M. Monier (France), B.-S. Moon and K.-W. Jun (Korea), A. Nijenhuis, D. E. Otero, P. P. Ray, Univ. of South Alabama Problem Group, Western Maryland College Problem Group, and the proposer.

Another Congruence Identity

E 3297 [1988, 953]. *Proposed by Gunnar Blom, University of Lund and Lund Institute of Technology, Sweden.*

If m and k are positive integers and $m + 2k = pr$, where $p \geq k$ is an odd prime and r is a positive integer, prove that

$$2^{m+1} \sum_{j=0}^{k-1} (-1)^j \binom{m+j}{j} \equiv 2^{r-1} \bmod p.$$

Solution by Peter Montgomery, Unisys and University of California at Los Angeles. If $j < p$, then $\binom{x}{j}$ is a polynomial in x with integer coefficients divided by a

number $j!$ that is relatively prime to p, which implies that the difference between $\binom{x}{j}$ and $\binom{p+x}{j}$ is divisible by p when x is an integer. If $0 \le j \le k-1$, then $j < p$, and we have

$$(-1)^j\binom{m+j}{j} = \binom{-1-m}{j} \equiv \binom{pr-1-m}{j} = \binom{2k-1}{j} \pmod{p}.$$

We can now evaluate the sum modulo p:

$$2^{m+1}\sum_{j=0}^{k-1}(-1)^j\binom{m+j}{j} \equiv 2^{m+1}\sum_{j=0}^{k-1}\binom{2k-1}{j} = 2^{m+1}2^{2k-1}/2$$

$$= 2^{m+2k-1} = 2^{pr-1} \equiv 2^{r-1} \bmod p.$$

Solved also by D. Callan, J. Duemmel, S. Gaignoux (France), J. J. Kostal (student), B. Lindström (Sweden), O. P. Lossers (The Netherlands), L. E. Mattics, A. Pedersen (Denmark), J. H. Steelman, N. Strauss (Brazil), and the editors.

A Characterization of the Least Common Multiple, Again

E 3300 [1988, 954]. *Proposed by H. Das Gupta, Kanchrapara, West Bengal, India.*

Let N be the set of natural numbers. Prove that there is exactly one mapping f from $N \times N$ to N having the following three properties for $x, y \in N$:

(i) $f(x, y) = f(y, x)$,

(ii) $f(x, x) = x$,

(iii) $(y - x)f(x, y) = yf(x, y - x)$ if $y > x$.

Solution I by Glenn Powers, Western Kentucky University, Bowling Green. The least common multiple satisfies the given properties. If f_1 and f_2 are two appropriate mappings but $f_1 \neq f_2$, then there exist $s, t \in N$ such that the product st is minimal and $f_1(s, t) \neq f_2(s, t)$. By property (ii), $s \neq t$, so we may assume that $t > s$. Now property (iii) implies that $f_1(s, t) = [tf_1(s, t - s)]/(t - s)$ and $f_2(s, t) = [tf_2(s, t - s)]/(t - s)$. This implies $f_1(s, t - s) \neq f_2(s, t - s)$, which contradicts our choice of s and t.

Solution II by J. B. Wilker, University of Toronto, Canada. The function must be the least common multiple. The functional equation (iii) can be written more suggestively as $f(x, y)/(xy) = f(x, y - x)/[x(y - x)]$ if $y > x$. Repeated applications of this, facilitated by (i) when $x > y$, transform the independent variables by the Euclidean algorithm and therefore reduce the expression itself to $f(d, d)\, d^{-2}$, where $d = \gcd(x, y)$. Property (ii) then implies that $f(d, d)d^{-2} = d^{-1}$ and hence that $f(x, y) = xyd^{-1} = \operatorname{lcm}(x, y)$.

Solution III by Duane W. Bailey, Amherst College, MA. It suffices to give an algorithm for explicitly computing each value of f. It is convenient to view these values as an infinite matrix $f(m, n)$. The diagonal $\{f(n, n)\}$ can be entered by property (ii). The formula $f(n, 1) = f(1, n) = f(1, n - 1)n/(n - 1)$ for $n > 1$ pro-

vides a recursive formula for entering the first row and column. If we now assume that the first $m-1$ rows and columns have been entered, then the formula $f(n, m) = f(m, n) = f(m, n-m)n/(n-m)$ for $n > m$ provides a recursive formula for computing the remaining values in the mth row and column.

Editorial comment. Except for a trivial change of notation, this problem is identical to E 2661 [1977, 487; 1978, 684]. This information came to the editors' attention through a remarkably late solution of E 2661 that arrived just recently. Almost all solvers either noted that the function was the least common multiple and established the result by induction, or rephrased the problem in terms of $g(x, y) = xy/f(x, y)$.

Solved also by the proposer and 71 other readers.

ADVANCED PROBLEMS

6625. *Proposed by Nicholas Strauss, Pontificia Universidade Católica do Rio de Janeiro, Brasil, and Jeffrey Shallit, Dartmouth College.*

If k is a positive integer, let $3^{v(k)}$ be the highest power of 3 dividing k. Put

$$r(n) = \sum_{i=0}^{n-1} \binom{2i}{i}$$

for positive integers n. Prove that

(i) $v(r(n)) \geqslant 2v(n)$,

(ii) $v(r(n)) = v\left(\binom{2n}{n}\right) + 2v(\mathrm{n})$.

6626. *Proposed by Bruce G. Dearden and Michael B. Gregory, University of North Dakota, Grand Forks.*

It is well known that if ϕ is a continuous function on $\mathbb{R}$, then the composition $\phi \circ f$ is Riemann integrable on $[a, b]$ whenever f is Riemann integrable on $[a, b]$. Is the converse true? That is, suppose ϕ: $\mathbb{R} \Rightarrow \mathbb{R}$ has the property that $\phi \circ f$ is Riemann integrable on $[a, b]$ whenever f is Riemann integrable on $[a, b]$. Does it follow that ϕ is continuous?

6627. *Proposed by G. Boese, Munich, West Germany.*

If n is a positive integer, put

$$\zeta_n(s) = \sum_{j=1}^{n} j^{-s}.$$

It is known that there are infinitely many n such that ζ_n has zeros in $\{s\colon \operatorname{Re} s \geqslant 1\}$. However, prove that

(i) if $n = 1, 2, 3, 4, 5$, then $\operatorname{Re} \zeta_n(s) > 0$ for $\operatorname{Re} s \geqslant 1$;

(ii) if $n = 6, 7, 8, 9, 10$, then $|\zeta_n(s)| > 0$ for $\operatorname{Re} s \geqslant 1$.

SOLUTIONS OF ADVANCED PROBLEMS

Gaps in the Sequence of Integers Expressible as Sums of Two Squares

6539 [1987, 303]. *Proposed by Claudia A. Spiro, SUNY at Buffalo.*

Let $b_1 < b_2 < b_3 < \cdots$ be the distinct positive integers expressible as sums of two squares of integers. Prove that for any given positive integer d the equality $b_{n+1} - b_n = d$ holds for infinitely many n.

Composite Solution by Douglas A. Hensley, Texas A & M University, College Station, and the proposer. We give a proof based on the following lemma, which is a consequence of the quadratic reciprocity law and Dirichlet's Theorem on primes in arithmetic progressions.

LEMMA. *If N is a positive integer which is not a square, there are infinitely many prime numbers p such that $p \equiv 3 \pmod 4$ and $\left(\frac{N}{p}\right) = -1$.*

Proof. Suppose $N = 2^{\alpha}Q$, where Q is odd.

If α is odd and $Q \equiv 3 \pmod 4$, any prime p such that $p \equiv 1 \pmod Q$ and $p \equiv 7 \pmod 8$ will do.

If α is odd and $Q \equiv 1 \pmod 4$, any prime p such that $p \equiv 1 \pmod Q$ and $p \equiv 3 \pmod 8$ will do.

If α is even and $Q \equiv 3 \pmod 4$, any prime p such that $p \equiv 1 \pmod Q$ and $p \equiv 3 \pmod 4$ will do.

If α is even and $Q \equiv 1 \pmod 4$, let q be a prime dividing Q to an odd multiplicity δ; then we may use any prime p such that $p \equiv 1 \pmod{Qq^{-\delta}}$, $p \equiv 3 \pmod 4$, and p is congruent to a particular quadratic non-residue modulo q.

In each case we have selected p to be in a certain coprime residue class modulo a divisor of $4N$; Dirichlet's theorem then guarantees that there are infinitely many such p. Thus the lemma is proved.

To prove the assertion of the problem, suppose first that d is odd. Put $D = (d-1)/2$, i.e., $d = 2D + 1$. We shall determine an arithmetic progression of numbers k such that

$$k^2 + D^2 + j \quad \text{is not a sum of two squares for } j = 1, 2, \ldots, 2D. \tag{1}$$

Then $k^2 + D^2$ and $k^2 + D^2 + 2D + 1$ are consecutive "sums of two squares" with difference $2D + 1 = d$.

Since $D^2 < D^2 + j < (D+1)^2$ for $1 \leqslant j \leqslant 2D$, we see that $D^2 + j$ is not a square and so we may apply the Lemma to find a prime p_j such that

$$p_j \equiv 3 \pmod 4, \qquad \left[\frac{-D^2 - j}{p_j}\right] = -\left[\frac{D^2 + j}{p_j}\right] = 1, \qquad p_j \neq p_1, p_2, \ldots, p_{j-1}$$

for $j = 1, 2, \ldots, 2D$. Thus there is a number β_j between 0 and $2p_j$ such that

$$\beta_j^2 \equiv -D^2 - j \pmod{p_j}, \qquad \beta_j^2 \not\equiv -D^2 - j \pmod{p_j^2} \qquad (j = 1, 2, \ldots, 2D).$$

(If the smallest positive solution of $x^2 \equiv -D^2 - j \pmod{p_i}$ does not satisfy $x^2 \not\equiv -D^2 - j^2 \pmod{p_j^2}$, merely add p_j to it.) Finally, if $k \equiv \beta_j \pmod{p_j^2}$ for $j = 1, 2, \ldots, 2D$, then $k^2 + D + j$ is exactly divisible by the first power of p_j and so is

not a sum of two squares ($j = 1, 2, \ldots, 2D$). Thus (1) holds and the case of odd d is settled.

The case of even d is similar but slightly more complicated in detail. Put $D = d/2 - 1$, i.e., $d = 2D + 2$. We shall determine an arithmetic progression of numbers k such that

$$k^2 + (k + D)^2 + j \quad \text{is not a sum of two squares for } j = 1, 2, \ldots, 2D + 1. \quad (2)$$

But then $k^2 + (k + D)^2$ and $k^2 + (k + D)^2 + 2D + 2 = (k - 1)^2 + (k + D + 1)^2$ are consecutive "sums of two squares" with difference $d = 2D + 2$.

In view of the identity $2(x^2 + y^2) = (x + y)^2 + (x - y)^2$, it suffices to choose k so that the number

$$2\{k^2 + (k + D)^2 + j\} = (2k + D)^2 + D^2 + 2j$$

is not a sum of two squares for $j = 1, 2, \ldots, 2D + 1$. Since $D^2 < D^2 + 2j < (D + 2)^2$ for $j = 1, 2, \ldots, 2D + 1$ and $D^2 + 2j \not\equiv (D + 1)^2 \pmod 2$, we see that $D^2 + 2j$ is not a square for $j = 1, 2, \ldots, 2D + 1$. Thus we may use the lemma to find a prime number p_j such that

$$p_j \equiv 3 \pmod 4, \qquad \left[\frac{-D^2 - 2j}{p_j}\right] = -\left[\frac{D^2 + 2j}{p_j}\right] = 1, \qquad p_j \neq p_1, p_2, \ldots, p_{j-1}$$

for $j = 1, 2, \ldots, 2D + 1$. Thus there is an integer γ_j between 0 and $4p_j$ such that

$$\gamma_j^2 \equiv -D^2 - 2j \pmod{p_j}, \qquad \gamma_j \equiv D \pmod 2, \qquad \gamma_j^2 \not\equiv -D^2 - 2j \pmod{p_j^2}$$

for $j = 1, 2, \ldots, 2D$. (If the smallest solution of

$$x^2 \equiv -D^2 - 2j \pmod{p_j}, \qquad x \equiv D \pmod 2$$

does not satisfy $x^2 \not\equiv -D^2 - 2j \pmod{p_j^2}$, merely add $2p_j$ to it.) Finally if $k \equiv (\gamma_j - D)/2 \pmod{p_j^2}$ for $j = 1, 2, \ldots, 2D + 1$, then

$$(2k + D)^2 + D^2 + 2j \equiv \gamma_j^2 + d^2 + 2j \pmod{p_j^2}$$

and so $(2k + D)^2 + D^2 + 2j$ will be divisible by p_j but not by p_j^2 ($j = 1, 2, \ldots, 2D + 1$). Thus $(2k + D)^2 + D^2 + 2j$ and hence $k^2 + (k + D)^2 + j$ is not a sum of two squares for $j = 1, 2, \ldots, 2D + 1$. Thus (2) holds and the case of even d is settled.

Solved also by L. E. Mattics.

Prescribed Zero Sets for an Entire Function and Its Derivatives

6570 [1988, 359]. *Proposed by L. A. Rubel, University of Illinois at Urbana-Champaign.*

(a) Let (z_n) and (z_n') be sequences in $\mathbb{C}$, neither with a finite limit point. Assume that if a complex number w occurs exactly k times in (z_n) where $k \geq 1$, then it

occurs exactly $k-1$ times in (z_n'). (Subject to this restriction, we allow finite or even empty sequences.) Show that there exists an entire function f such that the zeros of f are exactly at the points of (z_n) and the zeros of f' are exactly at the points of (z_n'), with the proper multiplicity in each case.

(b) Can one similarly prescribe three sequences (z_n), (z_n'), (z_n'') with corresponding assertions about the zeros of f, f', and f''?

Solution by the proposer. For convenience of notation, we will assume that the z_n are distinct—if they are not, then the proof below can be modified in a standard way. For a sequence (w_m) with no finite limit points, let $\Pi(z: w_m)$ denote a Weierstrass product with (w_m) as its zero set. We will repeatedly use the classical interpolation theorem that for any such (w_m) and a corresponding sequence (ξ_m) of complex numbers, there exists an entire function $A(z)$ such that $A(w_m) = \xi_m$ for all m.

In brief, the steps for the first part are the following:

Step 1. Choose an entire function $K(z)$ so that

$$K(z_n') = -\frac{\Pi'(z_n': z_m)}{\Pi(z_n': z_m)},$$

where Π' denotes the derivative of Π.

Step 2. Define

$$R(z) = \frac{K(z)\Pi(z: z_m) + \Pi'(z: z_m)}{\Pi(z: z_m')}.$$

Step 3. Choose an entire function $\delta(z)$ so that $\delta(z_n) = \log R(z_n)$, say for the principal branch of the logarithm.

Step 4. Define

$$M(z) = \frac{e^{\delta(z)} - R(z)}{\Pi(z: z_m)}.$$

Step 5. Define $h'(z) = K(z) + M(z)\Pi(z: z_m')$.

Step 6. Define $f(z) = e^{h(z)}\Pi(z: z_m)$. Then

$$\begin{aligned} f'(z) &= e^{h(z)}\left[h'(z)\Pi(z: z_m) + \Pi'(z: z_m)\right] \\ &= e^{h(z)}\left[\left\{K(z)\Pi(z: z_m) + \Pi'(z: z_m)\right\} + M(z)\Pi(z: z_m')\Pi(z: z_m)\right] \\ &= e^{h(z)}\Pi(z: z_m')\left[R(z) + e^{\delta(z)} - R(z)\right] = e^{h(z)+\delta(z)}\Pi(z: z_m'), \end{aligned}$$

and the result follows.

For the second part, the answer is *NO*, as we can see on trying to prescribe $Z(f)$, the zero set of f, by putting $Z(f) = Z(f'') = \varnothing$ and $Z(f') = \{0\}$. By [**HAY**, p. 67], we would have necessarily $f(z) = e^{az+b}$ from $Z(f) = Z(f'') = \varnothing$, but then $Z(f')$ would also be empty.

REFERENCE

[HAY] Walter K. Hayman, Meromorphic Functions, Clarendon Press, Oxford, 1964.

An Identity Involving Sums and Products

6575 [1988, 559]. *Proposed by Richard Stanley, Massachusetts Institute of Technology.*

Let $x_1, \ldots, x_n, y_1, \ldots, y_n$ be variables and for $i \leqslant j$ set $X_{ij} = x_i + x_{i+1} + \cdots + x_j$ and $Y_{ij} = y_i + y_{i+1} + \cdots + y_j$. Also set $X_{i+1,i} = Y_{i+1,i} = 0$, and any product $\prod_h^{h-1}$ equal to 1. Prove the identity

$$\sum_{i=1}^{n} \frac{\prod_{j=1}^{i} (X_{ji} + Y_{j,i-1}) \cdot \prod_{k=1}^{n+1-i} (X_{i+1,n+1-k} + Y_{i,n+1-k})}{\prod_{r=1}^{i-1} (X_{r+1,i} + Y_{r,i-1}) \cdot \prod_{s=1}^{n-i} (X_{i+1,n+1-s} + Y_{i,n-s})} = \sum_{i \leqslant j} x_i y_j.$$

For instance, when $n = 2$ we get

$$\frac{x_1(x_2 + y_1 + y_2)y_1}{x_2 + y_1} + \frac{(x_1 + x_2 + y_1)x_2 y_2}{x_2 + y_1} = x_1 y_1 + x_1 y_2 + x_2 y_2.$$

Solution by D. E. Knuth, Stanford University. Let

$$f(z) = -(x_1 + z) \prod_{j=1}^{n} \left(1 - y_j/(z - X_{2,j} - Y_{1,j-1})\right).$$

On a sufficiently large circle $|z| = R$, the Laurent expansion

$$\begin{aligned} f(z) &= -(x_1 + z) \prod_{j=1}^{n} \left(1 - y_j z^{-1} - y_j(X_{2,j} + Y_{1,j-1}) z^{-2} + O(R^{-3})\right) \\ &= -(x_1 + z)\left(1 - \frac{Y_{1,n}}{z} - \frac{\sum_{2 \leqslant j \leqslant k} x_j y_k}{z^2} + O(R^{-3})\right) \\ &= -z - x_1 + Y_{1n} + \sum_{j \leqslant k} x_j y_k z^{-1} + O(R^{-2}) \end{aligned}$$

converges uniformly. Integrating term by term gives the right side of the stated identity:

$$\frac{1}{2\pi i} \int_{|z|=R} f(z)\, dz = \sum_{j \leqslant k} x_j y_k.$$

This integral is also the sum of the residues inside; and the residue at $z = X_{2,k} + Y_{1,k-1}$ is

$$y_k(X_{1,k} + Y_{1,k-1}) \prod_{\substack{1 \leqslant j \leqslant n \\ j \neq k}} \frac{X_{2,k} + Y_{1,k-1} - X_{2,j} - Y_{1,j}}{X_{2,k} + Y_{1,k-1} - X_{2,j} - Y_{1,j-1}},$$

which is the kth term on the left side of the stated identity.

Editorial Comment. A more elementary but somewhat lengthier proof was provided by Sidney Heller (Brookhaven National Laboratory, Uptown, NY). The proposer informs us that this identity arose in the study of the irreducible characters of the symmetric group S_n. His proof involved the Young diagram and the

Frame-Robinson-Thrall "hook length formula." Knuth's solution above uses a more generally familiar technique, and has the additional virtue of exhibiting another (perhaps simpler) way of writing the left-hand side.

Professor Anatoly M. Vershik of Leningrad State University has informed us that the algebraic identity of this problem occurs as formula (12) of his paper [**1**]. He used essentially the same proof as the proposer, although his original motivation was entirely different.

1. A. M. Vershik, The hook formula and related identities (Russian), *Zap. Naučn Sem. Leningrad Otdel. Mat. Inst.* vol. 172, *Differential Geometry, Lie Groups, and Mechanics*, 10(1989) 3–20.

The solutions by Heller, Knuth, and the proposer were the only ones received.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

A Formalization of Set Theory without Variables. By Alfred Tarski and Steven Givant. American Mathematical Society, Colloquium Publications, Volume 41, Providence, 1987. xxi + 318 pp.

LISL GAAL
School of Mathematics, University of Minnesota, Minneapolis, MN 55455

Formalizations of the laws of thought and reasoning date back to Aristotle (384–322 B.C.), but set theory only dates back to the work of Georg Cantor (1845–1918) in the 19th century, yet the mathematical analysis of logic, set theory, and also modern algebra are now very closely intertwined and each benefits by advances in the other.

The language of mathematics, like all other languages, evolved gradually over the centuries and the notion of variables grew with it, almost assuming a life of its own. After all, what exactly do we mean when we start a sentence "For all $x, \ldots$" or say "There is an x such that $x^2 = -1$." What are all the x's we might have to consider? Variables are much like pronouns and so ought to be expendable. Omar Khayyam stated mathematical problems beautifully using poetry but no variables. Aside from the fact that few mathematicians are as adept at poetry, this method results in very convoluted linguistic tangles when applied to modern problems. The work of Boole and DeMorgan early in the 19th century not only showed that the union, intersection and complementation of sets follow specific algebraic laws, but also that sets can model the language and syllogisms of Aristotle. For example, if $A(x)$ and $B(x)$ are statements about some object x, and $A = \{\text{all } x\text{'s for which } A(x) \text{ is true}\}$ and $B = \{\text{all } x\text{'s for which } B(x) \text{ is true}\}$, then the sentence

$$\forall x(A(x) \wedge B(x)) \leftrightarrow (\forall x A(x) \wedge \forall x B(x))$$

is equivalent to the set-theoretic statement

$$A \cap B = 1 \Leftrightarrow (A = 1 \,\&\, B = 1).$$

The new statement does not contain any quantifiers '$\forall x$' nor '$\exists x$' and it is made up from equations only. Moreover it seems clear that there must be some relation between the Boolean algebra of sets and logical statements which works both ways: The Boolean sentence is true if and only if the corresponding logical statement is also true.

Cantor developed set theory as a basis for the natural number system and for a mathematical theory of the infinite, a subject heretofore considered untouchable. Although paradoxes were discovered in Cantor's version of set theory, later systems amended its shortcomings and Russell and Whitehead's monumental *Principia Mathematica* made set theory the basis of all mathematics. To carry out this task it is necessary to represent ordered pairs, n-tuples, relations and functions by sets: If we write '$\{x_1, x_2, \ldots\}$' to denote the set containing $x_1, x_2, \ldots$ then the ordered pair

$\langle x_1, x_2, \dots \rangle$ and n-tuple $\langle x_1, \dots, x_n \rangle$ can be defined by

$$\langle x_1, x_2 \rangle = \{\{x_1\}, \{x_1, x_2\}\},$$
$$\langle x_1, x_2, x_3 \rangle = \langle \langle x_1, x_2 \rangle, x_3 \rangle$$
$$\langle x_1, \dots, x_n \rangle = \langle \langle x_1, \dots, x_{n-1} \rangle, x_n \rangle.$$

A binary relation R between elements of a set A is then a set of ordered pairs $\langle x_1, x_2 \rangle$ whose elements are in the relation R: for example, the relation denoted by " $<$ " between integers consists of the set of all ordered pairs $\langle x_1, x_2 \rangle$ in which the integer x_1 is less than the integer x_2. Analogous definitions apply to relations on n elements. If R is a binary relation it is customary to write '$x_1 R x_2$' whenever $\langle x_1, x_2 \rangle$ belongs to the relation R. Functions are special kinds of relations, namely those for which any given element x_1 occurs in at most one pair in R. Numbers can also be thought of as sets of a special kind. For example, one can let

$$0 = \{\} \text{ (the empty set)},$$
$$1 = \{0\},$$
$$2 = \{0, 1\},$$
$$3 = \{0, 1, 2\},$$
$$\dots$$

and begin to build up the number system and mathematics from there.

The title of this book may lead one to believe that it concerns a special kind of set theory which needs no variables. This is however not so. Instead it is a pointer to one of the important theorems proved: Every one of the familiar axiom-sets for set theory, as well as those of a great many other axiom systems, can be equivalently formulated in a language $L^{\times}$ which has no variables.

In mathematics, variables are used primarily to serve as pronouns and to designate those places in a sentence which are all to be filled in with the name indicated in the corresponding quantifier "for all..." or "for some...." For example the rather awkward sentence "there is a real number which when cubed gives the same result as when it is multiplied by three" is customarily written

$$\exists x (x \text{ is a real number and } x^3 = 3x). \tag{1}$$

This makes the content of the sentence much clearer. But there are some difficulties with variables. For instance, when you say "for all x," do you really mean "for all objects in the universe" or "for all real objects in the universe and all abstract concepts that you, I or anyone else may ever conceive"? The latter presents a pretty tall order which may lead to some deep philosophical discussions. If it could be shown that quantifiers and variables are not really necessary one could avoid these deep (although often fascinating) problems. In this simple example it is easy to get rid of the variable x: Let Cu be the set of all pairs (r, r^3) of real numbers and let Tri be the set of all pairs $(r, 3r)$ of real numbers. These are now fixed sets, defined once and for all. Their meaning will not change with the context. Instead of (1) we can then write

$$\text{Cu} \cap \text{Tri} \neq \varnothing \tag{2}$$

(where $\varnothing$ is the empty set) and the (false) statement "every real number when multiplied by three is equal to its cube" becomes

$$\mathrm{Cu} = \mathrm{Tri}. \tag{3}$$

This sentence is now free of variables.

Even in only slightly more complicated cases, especially when there are several different variables in one sentence, more technical tricks are necessary to carry out their elimination and the resulting sentences become very contorted and difficult to handle. This is partly due to the fact that whenever several variables are involved one must use ordered n-tuples $\langle x_1, \ldots, x_n \rangle$ and their inverse functions $f_k(x_1, \ldots, x_k, \ldots, x_n) = x_k$ (for k fixed, $1 \leq k \leq n$). Since ordered n-tuples are easily definable in terms of iterated ordered pairs (i.e., $\langle x_1, \ldots, x_n \rangle = \langle \langle x_1, x_2, \ldots, x_{n-1} \rangle, x_n \rangle$), only the pairing function $Q\langle x, y \rangle = z$ and its inverses $A(z) = x$ and $B(z) = y$ need be present. (A very lucid informal exposition of the elimination of variables will be found in the article "Variables explained away" by Willard V. Quine (*Proc. of the Amer. Philos. Soc.*, 104 (1960) 343–347.) This book's result is, however, very much stronger than the one in this article.

The many known paradoxes of set theory have taught us to treat sets with kid gloves, so for safety's sake the first step must be the development of a suitable formal logic L. This is essentially a formalization of first order predicate logic with equality between individuals used throughout mathematics, except that only one constant binary predicate E is included, intuitively to correspond to the membership relation. The system L includes the sentential connectives $\wedge$ (and), $\vee$ (or), $\neg$ (not) and the quantifiers $\forall$ (for all) and $\exists$ (for some) as well as the customary laws and rules governing these. A set theory such as the Bernays-Gödel system can then be formulated in L by adding a suitable collection of axioms involving the predicate E. The simple examples above already suggest that it might be possible to avoid the use of variables and quantifiers by appropriate translation of the sentences of L into equivalent sentences about sets.

Again thinking of binary relations R, $S, \ldots$ as sets of ordered pairs and writing 'xRy' whenever the ordered pair $\langle x, y \rangle$ is a member of R, we then add the new symbols $+, -, \odot, ^{\cup}$ and $=$ to L and also the following axioms which tell us how they are used:

Let R, S be any binary relations, then

$$\forall x, y\{x(R + S)y \Leftrightarrow (xRy \vee xSy)\}$$

$$\forall x, y\{xR^{-}y \Leftrightarrow \neg(xRy)\}$$

$$\forall x, y\{x(R \odot S)y \Leftrightarrow \exists z(xRz \wedge zSy)\}$$

$$\forall x, y\{xR^{\cup}y \Leftrightarrow yRx\}$$

$$R = S \Leftrightarrow \forall x, y\{xRy \Leftrightarrow xSy\}.$$

Since these axioms are equivalences, the result is a definitional expansion of L, so it is not surprising that L and L^{+} are equivalent. By "equivalent" we mean that if a set U of elements is given and also a set $E \subseteq U \times U$, then U and E will satisfy all the axioms and theorems of L if and only if they also satisfy the axioms and theorems of L^{+}, provided $x, y, \ldots$ range over U and E is interpreted as E. The relation $=$ then becomes the equality of the elements of U. For convenience we let

$1 = \{\text{all pairs}\langle x, y\rangle\}$. The non-equality (2) is then equivalent to the equation

$$1 \odot (\mathrm{Cu} \cap \mathrm{Tri}) \odot 1 = 1, \tag{4}$$

since $1\ R\ 1 = 1$ whenever R is a non-empty set of pairs, that is whenever $\exists x, y\ \{xRy\}$. What's more, this equation does not contain any variables. The rule of substitution of equals for equals in equations of L^+ can be proved to hold, so the set of equations of L^+ forms a natural subsystem $L^\times$ of L^+ and is closed under the substitution rule. However even the simple statement that there exist more than three elements cannot be expressed in the form of an equation in L^+. This is a result of Korselt cited by L. Löwenheim (Über Möglichkeiten im Relativkalkül, *Math. Ann.*, 76 (1915) 470.) So $L^\times$ is not nearly as powerful as L^+. In fact it turns out that $L^\times$ is equivalent precisely to the fragment of L which has only three variables. Yet, despite the weakness of $L^\times$ and almost paradoxically in spite of Korselt's result, it is adequate for the formulation of the standard set theories and many others as well.

To state this more precisely, we call a system of axioms in L a Q-system if there are formulas in L which define a pairing function Q so that $Q(\langle x, y\rangle) = z$ is uniquely determined and also corresponding inverse functions A and B which allow us to recover x and y when z is given. The title theorem then states that every Q-system S in L is equivalent to some system S' in $L^\times$. But there is much more: For example, S' is decidable if and only if S is decidable, S' is finitely axiomatizable if and only if S is finitely axiomatizable. Since the standard set theories are Q-systems, these results can be applied. Many other systems are also Q-systems, most importantly the Peano Axioms, so they are also covered by these theorems. The preceding is perhaps an oversimplified account of the theorem which gives the book its title. There are several hundred other theorems in this volume and they are far from trivial. The book covers far more than just the standard set theories.

Algebraists began to study Boolean algebras in the 19th century, and it was also still in the 19th century that E. Schröder [Vorlesungen über die Algebra der Logik (exakte Logik), Leipzig, 1891] began the study of relation algebras. A relation algebra is a Boolean algebra whose elements are binary relations and with additional operations $\odot$, $\cup$, and a distinguished element $1'$ in A such that for all relations R, S, T in A we have

$$\begin{aligned}
(R \odot S) \odot T &= R \odot (S \odot T) \\
(R \cup S) \odot T &= (R \odot T) \cup (S \odot T) \\
R \odot 1' &= R \\
R^{\cup\cup} &= R \\
(R \cup S)^{\cup} &= R^{\cup} \cup S^{\cup} \\
(R \odot S)^{\cup} &= (R^{\cup} \odot S^{\cup}) \\
\left(R^{\cup} \odot (R \odot S)^{-}\right)^{-} \cup S &= 1.
\end{aligned}$$

These axioms are all algebraic in nature, in fact they are all equations. Models of set theories, as well as models of many other systems, are relation algebras provided the membership relation is among the elements of A. The equational axioms which characterize relation algebras are another example of the parallelism between logical and algebraic results, for one of the results proved here is the fact that these are the axioms of an undecidable theory and this result is essentially equivalent to Markov's

theorem [*Theory of Algorithms,* The Israel Program for Scientific Translations, Jerusalem, 1961] that the word problem for semigroups is unsolvable. The books by Henkin, Monk, and Tarski on cylindrical algebras [*Cylindrical Algebras,* North Holland Publishing Co., Amsterdam, Part I (1971), Part II (1985)] and Halmos' beautiful collection of papers on polyadic algebras [*Algebraic Logic,* Chelsea Publishing Co., New York, 1962] are more evidence of the close relationship between algebra and logic. Together with Tarski-Givant's book they also provide an excellent bibliography of algebraic logic. Tarski-Givant is the newest of these books and very likely not the last, if the many open problems stated throughout are any indication. It is not a book for the casual reader. It requires a lot of attention to detail and technicalities, as well as familiarity with related literature to appreciate the results. Many of these will probably lead to further important applications.

Complex Analysis. By Joseph Bak and Donald J. Newman. Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1982, x + 244 pp.

Invitation to Complex Analysis. By Ralph Philip Boas. The Random House/Birkhäuser Mathematics Series, Random House, New York, 1987, xiv + 348 pp.

LAWRENCE ZALCMAN
Department of Mathematics and Computer Science, Bar-Ilan University, 52900, Ramat-Gan, Israel

By and large, mathematical disciplines get the texts they deserve. Certainly this has been the case with complex function theory. That "showpiece of nineteenth century mathematics" has spawned some of the great books of twentieth century analysis—a short list would include Hermann Weyl's *The Concept of a Riemann Surface*, Pólya and Szegö's *Problems and Theorems in Analysis*, and Nevanlinna's *Analytic Functions*, classics all, fresh as the day they were written, fifty, sixty, seventy years ago and more—as well as a host of superb graduate texts. Indeed, the instructor in a first graduate course in complex variables faces a genuine embarrassment of riches: a selection of treatments distinguished not only for their mastery of the subject matter but also for the freshness and distinctiveness of their point of view. One need only mention, among the older texts, such masterpieces as Carathéodory's *Theory of Functions* and *Analytic Functions* by Saks and Zygmund. The current market seems to be dominated, deservedly, by Ahlfors' authoritative *Complex Analysis* and Rudin's extremely elegant *Real and Complex Analysis*. Nor should one pass in silence over such valuable recent additions to the literature as Narasimhan's beautiful *Complex Analysis in One Variable* or the late Peter Henrici's three-volume *Applied and Computational Complex Analysis*, which history may well mark as a *meisterwerk* of enduring value.

Alas, so far as *undergraduate* texts are concerned, the situation is very different indeed. Here a kind of Gresham's Law for textbooks applies, with stale texts driving out fresh—on those rare occasions that a fresh one manages to get written. The result (to exaggerate only slightly) has been a drab and colorless phalanx of Churchill clones, presenting the same material in the same way in the same order.

All that is changed now with the appearance of the books under review. Written by men who really understand complex variables, they are, quite simply, better than anything else in the field; and if, despite somewhat varying aims, they seem remarkably similar in their choice of material (even beyond the canonical core of a first course in complex analysis), that is simply a reflection of a shared assessment of what is genuinely important in complex function theory.

The stated aim of *Complex Analysis* by Bak and Newman (based on lectures delivered by the second author at Yeshiva University in the early seventies) is "to present the theory of analytic functions with as little dependence as possible on advanced concepts from topology and several-variable calculus... also to highlight the authentic complex variable methods and arguments as opposed to those of other mathematical areas." That aim is realized splendidly. While there are certainly novelties of exposition (of which the development of the properties of entire functions before a detailed study of the local theory is perhaps the most notable), what is really exciting about this text is the nonstandard material it contains. This applies even, and especially, to the discussion of familiar results. Thus Liouville's Theorem is given not only in its usual version ("a bounded entire function is constant") but also in extended form, in which an inequality of the form $|f(z)| \leqslant A + B|z|^k$, or even a one-sided inequality like $\operatorname{Re} f(z) \leqslant A|z|^k$ for sufficiently large z, implies that f is a polynomial of degree at most k. (These results are also presented in Boas. For an alternate development, available quite early in a first course, see [**12**].) A related result, less likely to be familiar to the reader, is that an entire function which satisfies $|f(z)| \leqslant 1/|\operatorname{Im} z|$ must vanish identically. The maximum modulus principle is accompanied by an "anti-calculus proposition," attributed to Erdös, according to which a nonconstant function f analytic on a closed disc *cannot* take its maximum at a boundary point at which f' vanishes. (Boas also states this proposition, which he, citing Pólya-Szegö, attributes to Loewner.) Newman's treatment of sums and integrals via residue methods ranges beyond the usual examples to a discussion of the evaluation of *indefinite* integrals of rational functions (also discussed by Boas) as well as the calculation and (in general, much more difficult) estimation of sums involving binomial coefficients. It is shown, for instance, that

$$\lim_{n\to\infty} \sum_{k=0}^{n} (-1)^k \sqrt{\binom{n}{k}} = 0;$$

try proving that without complex variables.

In addition to the standard material of a first course through the Riemann mapping theorem and Poisson's formula, various special topics are developed. Particularly noteworthy is the discussion of theorems of Phragmén-Lindelöf type, a topic absent even in Ahlfors. These results are used to derive Iversen's Theorem (a nonconstant entire function tends to infinity along some path tending toward infinity) as well as to prove that a solution of $f'(z) = -f(z/2)$ analytic at 0 must be an entire function unbounded on every ray. The final chapter, devoted to *mathematical* applications of function theory, shows how complex variable methods can be applied to partition problems and the study of infinite systems of equations and even used to calculate the total variation of $(\sin x/x)^2$ on the line. Also

included here is a very slick proof of the uniqueness theorem for Fourier transforms on the line. The book ends on a high note with Newman's own streamlined proof of the Prime Number Theorem. *Complex Analysis* by Bak and Newman is a book that lives and breathes the excitement of complex variables.

Ralph Boas' *Invitation to Complex Variables* is every bit as good, and better. The distillation of an intimate acquaintance with the field that extends over half a century, it is, in my opinion, the best undergraduate text ever written on the subject. The years of thought that must have gone into the making of this book are evident not simply in its organization and choice of topics (both of which are exemplary) but also in many little touches. Boas is the only author I know who approaches Morera's theorem from the proper direction (viz. Green's formula); and even when he comes to prove a result as familiar as the fundamental theorem of algebra, he has something interesting to show us. Again and again, one encounters illuminating discussions of matters ignored entirely by most other authors. Cases in point include the section on changing variables in improper integrals and the discussion of methods for calculating harmonic conjugates. Another nice feature of the book are the notes, which contain well-chosen references to the literature, attributions, pointed remarks, an occasional reminiscence (note 1, p. 116), and some amusing speculations (note 2, p. 116; note 2, p. 225).

As Boas observes early on, "complex analysis was originally developed for the sake of its applications." In addition to the usual applications to irrotational flows of incompressible, nonviscous fluids, there are discussions of the Plemelj formulas, Nyquist diagrams, and asymptotic series. But make no mistake about it: this is a *mathematics* book. The reader will encounter here the Vivanti-Pringsheim theorem, the Hadamard gap theorem, and a particularly detailed discussion of Phragmén-Lindelöf theorems (all of which are also in Bak-Newman) as well as Tauber's theorem, Vitali's theorem, Jensen's formula, and the elementary theory of univalent functions (which are not).

Indeed, there is so much material here beyond what is contained in the usual undergraduate semester of complex variables that it makes more sense to ask what is missing. The general (homology) form of Cauchy's theorem, for one thing. Quite rightly, neither Boas nor Newman believes in getting bogged down in topology in a first course in function theory. (As to the possibilities in this direction, see [**5**], where 45 pages are devoted to the proof of Cauchy's theorem!)[1] For similar reasons, the monodromy theorem is not so much as mentioned in either text. Less easy to understand or, in view of the simplicity of the proof, to justify is the absence of the Mittag-Leffler theorem on prescribing principal parts of meromorphic functions. I was also surprised not to find the *symmetric* version of Rouché's theorem ("if f and g are analytic on and inside the simple closed curve Γ, and $|f + g| < |f| + |g|$ on Γ, then f and g have the same number of zeroes inside Γ"); true, no other text of my acquaintance includes it either, but somehow one comes to expect more from Boas and Newman. Both authors discuss infinite products, prove Wallis' formula and

[1] Of course, there are various shortcuts; and Dixon's elegant approach [**3**], developed *in extenso* in [**9**, pp. 217–220], removes the last excuse for not including this material in a first graduate course. Still, I would argue that Cauchy's theorem in its most general form is precisely the sort of thing one *should* leave out in an undergraduate course.

derive the product representation for $\sin \pi z$. Newman proves the Weierstrass product theorem; Boas, remarking that the Weierstrass representation "is so general that it is not very useful," merely states it but then compensates the reader by also stating (again without proof) the Hadamard factorization theorem. Boas also omits any mention of the Riemann zeta function. This last omission, I think, was a mistake: even in the absence of any application to the Prime Number Theorem, discussion of the properties of the zeta function offers an ideal opportunity to illustrate many of the basic ideas of function theory in a setting that is of special interest. But by far the most serious omission is Picard's Theorem, which is stated in passing in both texts but otherwise ignored. Actually, in view of [**12**], there is no reason not to include (a version of) this high point of complex analysis even in a course for beginners. One imagines that in subsequent editions (of which there should be many) some of these gaps will be filled.

Exercises are a strong point in both of these texts. They are well-chosen and advance the student's understanding of the material. In addition, Boas has a unique feature: 75 pages of complete solutions to the exercises. This enables him to leave certain more-or-less routine verifications (as well as further developments) as exercises and makes the book particularly suitable for self-study. Also included are some supplementary exercises, for which solutions are not given.

Both volumes are nicely produced (as one would expect) and excellent value for the money. A cursory reading turned up only a couple of slips in Bak-Newman: on page 88, line 9, read "set" for "domain"; and on page 196, lines 2 and 3, read m for k. Boas contains a number of misprints, a (partial) list of which is available from the author. Most are harmless, but the example used to illustrate the definition of line integral on page 36 contains an error likely to confuse beginners. One should also omit the last complete sentence on page 171 as well as the phrase "with $\varepsilon > 0$" on page 246, line 4. In note 1 on page 172, read 25E for 26E; and on page 219, line 7, change "two" to "one." There is a genuine (though inessential) slip at the bottom of page 257, where "nonvanishing" should be inserted before "functions" on line 6; note 1 on page 261 needs corresponding modification (or deletion). The solution to exercise 35.2(a) is incorrect.

Such minor blemishes should not be given undue weight. The books here reviewed are the texts of choice for an undergraduate course (or seminar) in complex analysis for mathematics majors. They belong on the bookshelf of every analyst and, for that matter, of anyone who ever plans on teaching a course in function theory—at any level. An even better place for them is in the hands of the student who wants to know what complex variables is really all about.

REFERENCES

1. Lars V. Ahlfors, Complex Analysis, 3rd ed., McGraw-Hill, 1979.
2. C. Carathéodory, Theory of Functions of a Complex Variable, vols. I and II, 2nd English ed., Chelsea, 1954 (translation of Funktionentheorie, Birkäuser, 1950).
3. John D. Dixon, A brief proof of Cauchy's integral theorem, *Proc. Amer. Math. Soc.*, 29 (1971) 625–626.
4. Peter Henrici, Applied and Computational Complex Analysis, vols. 1, 2, and 3, Wiley, 1974, 1977, and 1986.
5. Kunihiko Kodaira, Introduction to Complex Analysis, Cambridge University Press, 1984 (translation from Japanese of Complex Analysis, vols. I and II, Iwanami Shoten, 1977 and 1978).

6. Raghavan Narasimhan, Complex Analysis in One Variable, Birkhäuser, 1985.
7. Rolf Nevanlinna, Analytic Functions, Springer-Verlag, 1970 (revised translation of Eindeutige analytische Funktionen, 2nd ed., 1953; 1st ed., 1936).
8. G. Pólya and G. Szegö, Problems and Theorems in Analysis, vols. I and II, Springer-Verlag, 1972 and 1976 (revised and enlarged translation of Aufgaben und Lehrsätze aus der Analysis, 4th ed., 1970 and 1971; 1st ed., 1924).
9. Walter Rudin, Real and Complex Analysis, 3rd ed., McGraw-Hill, 1986.
10. Stanislaw Saks and Antoni Zygmund, Analytic Functions, 3rd ed., Elsevier, 1971 (translation of Funkcje Analityczne, Monografie Matematyczne, 1938).
11. Hermann Weyl, The Concept of a Riemann Surface, 3rd ed., Addison Wesley, 1955 (revised translation of Die Idee der Riemannschen Flächen, Teubner, 1913).
12. Lawrence Zalcman, Picard's theorem without tears, *Amer. Math. Monthly*, 85 (1978) 265–268.

Fractals Everywhere. By Michael Barnsley. Academic Press, Boston, 1988. xii + 394 pp.

ANTHONY BARCELLOS
Department of Mathematics, American River College, Sacramento, CA 95841

> I am very ambivalent about fractals becoming a field because my work was inspired and supported by a very strong belief that the division of science into fields has been extremely harmful.
>
> Benoit B. Mandelbrot, 1983 [**1**]

Occasionally it is possible to coax an opinion out of Benoit Mandelbrot. As proud as the "Father of Fractals" may be of his brainchild, he looks toward its domestication with trepidation. Known as a polymath, Mandelbrot has carried his fractal message anywhere that people might listen, collecting academic appointments in mathematics, applied mathematics, economics, engineering, physiology, and medicine [**2**]. His success has been singular.

Fractal images have become a commonplace of math department offices and computer centers. Fractal techniques have become acceptable in cosmology, dynamical systems, and computer graphics. Not only is fractal geometry becoming a respectable tool, it is, despite Mandelbrot's concerns, becoming a topic in its own right.

The evidence of this is everywhere (along with the fractals themselves). Since the last edition of Mandelbrot's original fractal manifesto [**6**], we have been treated to lavish picture books [**7**], best-selling popularizations [**5**], meticulous treatises with yet more pictures [**8**], and philosophical collaborations between physicists and psychologists [**4**].

Among these varied offerings we have the present book, *Fractals Everywhere*, by Michael Barnsley of the Georgia Institute of Technology and Iterated Systems, Inc. It is a textbook for a course on fractals. However, Benoit need not fear that his creation is being set up for codification and sterile abstraction. Barnsley treats fractals as the unifying thread in his presentation of metric spaces, mappings, dynamical systems, image compression [**3**], and graphical models.

Fractals Everywhere is yet another picture book. This is good, because anyone who ignores the visual side of fractals abandons not only its esthetic dimension, but also its ability to represent what we otherwise cannot see. Mandelbrot's *The Fractal Geometry of Nature* is in part a casebook of missed opportunities, where he points out how inadequate visualization of complex systems obscures their nature (see [**6**, p. 179] in particular).

Barnsley's cover illustration, a depiction of an Andes Indian girl, is a collage of 160 iterated maps. The emphasis on visuals is a fundamental part of the book and not, as in so many other cases, a sugar-coating with little relation to what's inside. Barnsley's notation is often iconic; that is, specialized symbols provide strong visual reminders of what they represent. Given the inherent complexity of the material, it's nice to see a heavy dot representing a closed disk in $\mathbf{R}^2$ and a filled square standing in for the unit square; that's easier than having to remember what D and S stand for (or whatever). Even more striking are the icon for a Sierpiński triangle and the little mummy representing a frozen cadaver (a useful $\mathbf{R}^3$ subset that Barnsley calls "Body space").

These playful elements pervade the book, and a good thing, too. Barnsley uses them as memory devices to minimize the mental baggage one must carry during his fractal trek, leaving more room for the booty to be gathered as the trip progresses.

Barnsley's fractal geometry course at Georgia Tech was the basis for his text. The prerequisite is two years of calculus, which makes the entry requirements rather low; a healthy dollop of mathematical maturity would be a good idea. Ramping up from an elementary discourse on the properties and topology of metric spaces, Barnsley moves into contraction mappings and defines the iterated function system (IFS), a tool that recurs time and again in all that follows. The probabilistic Chaos Game is contrasted with the Deterministic Algorithm for fractal generation. Barnsley introduces dynamical systems and fractal dimension.

By this time the atmosphere is getting rarefied as Barnsley tracks through fractal interpolation, Julia sets, and measures on fractals. In actual practice, Barnsley reserves the latter chapters of his text for special topics that round out the basic introductory material. The rich selection ensures that each instructor will be able to tailor his or her own course in many different ways.

Fractals Everywhere has many praiseworthy features. The prose is clear and generally straightforward, though with enough passive voice so we don't forget we're in a textbook. Exercises abound—at all levels of difficulty. Practically every page has at least one illustration, and two-page spreads of pure text are almost impossible to find. (In fact, given Barnsley's predilection for iconic notation, several of the text pages could be said to be illustrated as well.) However, some of the figures are rather crudely drawn, and they suffer by comparison with the slick computer graphics that dominate the illustrations.

The author occasionally forgets that his readers may not be as familiar with fractal terminology as he is. Thus he mentions the Chaos Game or Chaos Algorithm in a few places, but spends a lot more time talking instead about the Random Iteration Algorithm. They're one and the same, but I never saw where Barnsley expressly said so. Unfortunately, the index references for the two names are disjoint, with no cross-references to each other.

The book may claim to have 9 chapters, but it really has only 8. I regret that the author refers to his five-page introduction as Chapter 1, as it is not even a member

of the same species as the other chapters. The concept of Chapter 0 was invented for chapters of measure zero like this one. This is, of course, a trifling matter, but so is Chapter 1.

We can probably expect that fractal geometry will become a fixture in the college mathematics curriculum. As a descriptive tool it has great power, providing a new model for natural processes that has already proved fruitful in many fields. *Fractals Everywhere* takes Mandelbrot's thesis of fractal ubiquity and frames it with sufficient structure to make classroom presentation possible. Barnsley's text will inspire rival books, but I think that *Fractals Everywhere* will be commonly referred to as the "classic" in its field.

Perhaps, somewhere, someone is writing (or has written) a textbook on fractals that begins with a definition in terms of Hausdorff-Besicovitch dimension and proceeds to adduce therefrom a numbered sequence of theorems and corollaries. This is not that book. Benoit would be pleased.

REFERENCES

1. Anthony Barcellos, The fractal geometry of Mandelbrot, *College Mathematics Journal*, 15 (March 1984).
2. —, Interview with Benoit Mandelbrot, Mathematical People, Donald J. Albers and G. L. Alexanderson, editors, Birkhäuser Boston, Cambridge, 1985.
3. Michael F. Barnsley and Alan D. Sloan, A better way to compress images, BYTE, 13 (January 1988) 215–223.
4. John Briggs and F. David Peat, Turbulent Mirror, Harper & Row, New York, 1989.
5. James Gleick, Chaos: Making a New Science, Viking, New York, 1987.
6. Benoit B. Mandelbrot, The Fractal Geometry of Nature, W. H. Freeman, New York, 1983.
7. Heinz-Otto Peitgen and Peter H. Richter, The Beauty of Fractals, Springer-Verlag, New York, 1986.
8. Heinz-Otto Peitgen and Dietmar Saupe, editors, The Science of Fractal Images, Springer-Verlag, New York, 1988.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S, L*.** *The World of Mathematics.* James R. Newman. Microsoft Pr, 1988, $13.95 (P) each [ISBN: 1-55615-148-9] set. *Volume One,* xxx + 711 pp; *Volume Two,* ix + 677 pp; *Volume Three,* viii + 602 pp; *Volume Four,* viii + 449 pp. A "faithful reproduction" of Newman's 1956 masterful anthology of mathematics, scanned and retypeset with the aid of modern computers. In this edition an enlarged 60-page comprehensive index is printed in each volume. Opens with a new Foreword by Philip and Phyllis Morrison. A superb resource for school and college libraries and for any teacher of mathematics. Perhaps its republication will inspire a worthy sequel focused on contemporary mathematics. LAS

Precalculus, T(13). *Precalculus: Functions and Graphs, Fifth Edition.* M.A. Munem, J.P. Yizze. Worth, 1990, xii + 534 pp, $38.95. [ISBN: 0-8791-418-0] Standard precalculus text, including polynomial, rational, exponential, logarithmic, and trigonometric functions as well as systems of equations and inequalities, conic sections and analytic geometry, sequences and series, and mathematical induction. Occasional diagrams are poorly drawn or printed. (*Fourth Edition,* TR, November 1986.) FR

Education, P, L*. *Mathematicians and Education Reform.* Eds: Naomi Fisher, Harvey Keynes, Philip Wagreich. CBMS Issues in Math. Educ., V. 1. AMS and MAA, 1990, viii + 217 pp, (P). [ISBN: 0-8218-3500-9] First in a new series of CBMS reports on mathematics education. This volume—the proceedings of a July 1988 workshop—contains papers on projects (by Tom Berger, Frank Demana, Uri Treisman, Herb Clemens, R.O. Wells, *et al.*) and short notes on issues (recruitment of women, nontraditional instruction, etc.). Focus is on contexts that enable university mathematicians to work effectively in mathematics education. LAS

Linear Algebra, T(13: 1), S, C, L. *Matrices.* North Carolina School of Science and Mathematics. NCTM, 1988, vi + 115 pp, $10 (P). [ISBN: 0-87353-270-8] A unit designed to show how matrices can be used to model real-world phenomena and as discrete data structures. The first three sections introduce matrix arithmetic and then studies the Leontief model, the Leslie model, Markov chains, and computer graphics. Well-written with lots of good examples and exercises. Includes easy-to-use software that allows static and dynamic matrices of up to 15 by 15 and is able to perform standard matrix computations. CEC

Group Theory, P. *The Classical Groups and K-Theory.* Alexander J. Hahn, O. Timothy O'Meara. Grund. der math. Wissenschaften, B. 291. Springer-Verlag, 1989, xv + 576 pp, $119. [ISBN: 0-387-17758-2] Covers algebraic properties of the classical groups over rings. Presents the development of topics which have evolved since Dieudonné's *La Géometrie des Groupes Classiques.* In two parts: the first covers the simple situation of the linear groups, and the second then does a parallel study for the other classical groups. Contains several new results. RH

Algebra, T(16: 1), S, P*, L. *Handbook of Boolean Algebras, Volume 1-3.* Ed: J. Donald Monk, Robert Bonnet. North-Holland (US Distr: Elsevier Science), 1989, $309.75 set [ISBN: 0-444-87291-4]. *Volume 1,* xix + 323 pp, $68.25 [ISBN: 0-444-70261-X]; *Volume 2,* xix + 415 pp, $112.25 [ISBN: 0-444-87152-7]; *Volume 3,* xix + 650 pp, $161. [ISBN: 0-444-87153-5] Treats those areas of the theory of Boolean algebras of most interest to pure mathematicians, i.e., set-theoretical abstract theory, applications, and relationships to measure theory, topology, and logic. *Volume 1* is a self-contained treatment of the subject including exercises. *Volumes 2* and *3* consist of articles that are more or less independent of each other which expose most of the frontiers of research in the subject. CEC

Algebra, T(18: 1, 2). *Commutative Ring Theory.* Hideyuki Matsumura. Transl: M. Reid. Stud. in Adv. Math., V. 8. Cambridge U Pr, 1989, xiii + 320 pp, $24.95 (P). [ISBN: 0-521-36764-6] Changes from the *First Edition* (TR, November 1987) include correction of mistakes, improvements of some theorems, and additional topics added in the form of appendices to individual sections. The section entitled "Kunz' Theorems" in the *First Edition* has been re-

placed, and the bibliography has been enlarged. RH

Calculus, T(13-14: 1). *Brief Calculus with Applications.* Raymond F. Coughlin, David E. Zitarelli. Saunders College, 1990, xvi + 476 pp. [ISBN: 0-03-031597-2] A one-semester calculus text for business, social, and biological science students. Indexed applications, gender-neutral problems, clear graphs, readable text. Includes referenced problems to support student projects or in-depth study. FR

Calculus, T(13-14: 2). *Applied Calculus, Second Edition.* Dennis D. Berkey. Saunders College, 1990, xix + 681 pp, $38. [ISBN: 0-03-031267-1] First year calculus text for business, economics, and social science students. Indexed applications, gender-neutral problems, clear graphs, readable text. BASIC applications programs in appendix. Includes infinite series, differential equations, probability theory. FR

Numerical Analysis, P. *Lecture Notes in Mathematics-1386: Numerical Methods for Ordinary Differential Equations.* Eds: A. Bellen, C.W. Gear, E. Russo. Springer-Verlag, 1989, vii + 136 pp, $13.50 (P). [ISBN: 0-387-51478-3] Proceedings of a workshop held in L'Aquila, Italy, September 16-18, 1987. Eight papers describe recent work on the development of improved general-purpose codes and algorithms for parallel computers. AO

Numerical Analysis, T(16-18: 1, 2). *Matrix Computations, Second Edition.* Gene H. Golub, Charles F. Van Loan. Ser. in Math. Sci., V. 3. Johns Hopkins U Pr, 1989, xix + 642 pp, $59.50; $29.95 (P). [ISBN: 0-8018-3772-3; 0-8018-3739-1] A revised and expanded version of this important text and reference. New material has been added throughout and some stylistic improvements have been made. The most significant change is the addition of new material (including a new chapter) on parallel matrix computation. (*First Edition,* TR, January 1985.) AO

Numerical Analysis, S(15-16). *Guide to Numerical Analysis.* Peter R. Turner. Math. Guides. Macmillan Education (US Distr: Scholium Intern), 1989, x + 208 pp, $22.50 (P). [ISBN: 0-333-44947-9] A brief introduction to many of the ideas and techniques discussed in an introductory numerical analysis course together with some of the relevant background material from first-year calculus. Designed to be used as a supplementary source of information for students taking a numerical analysis course. AO

Functional Analysis, P. *Nonlinear Functional Analysis and its Applications.* Eberhard Zeidler. Transl: Eberhard Zeidler, Leo F. Boron. Springer-Verlag, 1990. *II/A: Linear Monotone Operators,* xviii + 467 pp, $124 [ISBN: 0-387-96802-4]; *II/B: Nonlinear Monotone Operators,* xv + 733 pp, $115. [ISBN: 0-387-97167-X] The second of a five-volume series on nonlinear functional analysis. The theory of monotone operators generalizes Hilbert-space methods (for linear partial differential equations) to nonlinear problems. The first subvolume (*II/A*) is an introduction to the functional analytic theory of linear partial differential equations; the second (*II/B*) deals with the generalizations. (*II/B* may be read independently.) MLR

Analysis, P. *Thirteen Papers Translated from the Russian.* S.D. Berman, *et al.* AMS Transl. Ser. 2, V. 144. AMS, 1989, viii + 140 pp, $58. [ISBN: 0-8218-3125-9]

Algebraic Geometry, P. *Singularities.* Ed: Richard Randell. Contemp. Math., V. 90. AMS, 1989, xv + 359 pp, $42 (P). [ISBN: 0-8218-5096-2] The proceedings of an international conference on algebraic geometry held in July 1986 at the University of Iowa. Includes 19 papers on singularities in algebraic and analytic varieties, together with a report on a session devoted to problems and open questions. PZ

Differential Geometry, P. *Lecture Notes in Mathematics-1410: Differential Geometry.* Eds: F.J. Carreras, O. Gil-Medrano, A.M. Naveira. Springer-Verlag, 1989, 308 pp, $28.60 (P). [ISBN: 0-387-51885-1] Proceedings of the Third International Symposium held in Peñiscola, Spain, June 1988. MPR

Operations Research, T(15-18: 1, 2), P, L. *The Theory of Choice and Decision Making.* Igor M. Makarov, *et al.* MIR (US Distr: Imported Pub), 1987, 328 pp, $13.95. [ISBN: 0-8285-4017-9] A well-written translation from the Russian. The text is organized into three parts: the theoretical background for the choice of alternatives, procedures and algorithms for decision making, and optimal control with multiple criteria. The book is suitable for advanced undergraduates studying mathematics. Extensive bibliography, although about two-thirds of the references are in Russian. There are exercises scattered throughout the text, most rather difficult. The price of this book is extremely reasonable. SM

Operations Research, T(15-16), L. *Readings in Decision Analysis.* Simon French. Chapman & Hall, 1989, ix + 210 pp, $25 (P). [ISBN: 0-412-32170-X] This is an introduction to decision analysis. The author uses a lively combination of his own narrative and previously published papers to lead the reader through the theory and applications in the field. Reprints of six case studies are included. The mathematical sophistication required is minimal: the most intimidating symbols are summation signs and functions denoted by Greek letters. No exercises. SM

Operations Research, T(17-18). *Multiple Criteria Optimization: Theory, Computation, and Application.* Ralph E. Steuer. Robert E Krieger, 1989, xx + 546 pp, $57.95. [ISBN: 0-89464-393-2] Written as a text for graduate students in management, this book requires modest mathematical background (linear algebra with linear programming, use of calculus for optimization, willingness to use computer software). Good exposition that stays close to geometric representation throughout. Reprint of 1986 Wiley edition (TR, November 1987). AWR

Optimization, P. *Simulated Annealing (SA) & Optimization: Modern Algorithms with VLSI, Optimal Design, & Missile Defense Applications.* Ed: Mark E. Johnson. Amer. J. of Math. & Management Sci., V. 8, Nos. 3 & 4. American Sciences Pr, 1988,

245 pp, $89.75 (P). [ISBN: 0-935950-18-4] An annotated bibliography and six papers. Three of the papers describe applications in the areas of VLSI, optimal design, and missile defense. The other three papers discuss computational aspects of the technique. AO

Optimization, T(18: 2), P, L. *Methods of Dynamic and Nonsmooth Optimization.* Frank H. Clarke. CBMS-NSF Reg. Conf. Ser. in Appl. Math., V. 57. SIAM, 1989, v + 90 pp, $15.75 (P). [ISBN: 0-89871-241-6] A short and succinct monograph. Beginning with a chapter on nonsmooth analysis and geometry, the text continues through nonsmooth problems in the calculus of variations, verification functions, dynamic programming, and optimal control. The book ends with an index and a substantial list of references. No exercises; inexpensive. SM

Control Theory, P. *Lecture Notes in Control and Information Sciences-123: Knowledge-Based Control with Application to Robots.* C.W. de Silva, A.G.J. MacFarlane. Springer-Verlag, 1989, x + 196 pp, $30.30 (P). [ISBN: 0-387-51143-1] A research-level monograph describing recent work using machine intelligence techniques in the design of robot control mechanisms. AO

Control Theory, P, L. *Lecture Notes in Control and Information Sciences-124: Control of Partially-Known Dynamical Systems.* A.A. Bahnasawi, M.S. Mahmoud. Springer-Verlag, 1989, xi + 228 pp, $30.30 (P). [ISBN: 0-387-51144-X] The research in this monograph is primarily focused on reduced-order adaptive control schemes and deterministic stabilizing control for uncertain dynamical systems. Both discrete systems and continuous systems are discussed. SM

Control Theory, P, L. *Modern Optimal Control.* Ed: Emilio O. Roxin. Lect. Notes in Pure & Appl. Math., V. 118. Marcel Dekker, 1989, xiii + 437 pp, (P). [ISBN: 0-8247-8168-6] A collection of twenty-nine papers offering both an historical perspective on thirty years of developments in the study of optimal control and dynamical systems, and significant examples in modern control theory. An index of topics is provided. SM

Control Theory, T(18), P, L. *Lecture Notes in Control and Information Sciences-138: Robust Controller Design Using Normalized Coprime Factor Plant Descriptions.* D.C. McFarlane, K. Glover. Springer-Verlag, 1990, x + 206 pp, $30.30 (P). [ISBN: 0-387-51851-7] This book discusses the *robust stabilization problem* which involves finding a single feedback controller for all members of a model which are close, in some sense, to a nominal model. The motivation for studying this problem is the realization that modeling is an inexact art and parameter values are seldom known precisely. A robust controller will stabilize any physical system which is close enough to the system represented by a mathematical model. SM

Systems Theory, S(18), P. *Lecture Notes in Control and Information Sciences-135: Three Decades of Mathematical System Theory.* Eds: H. Nijmeijer, J.M. Schumacher. Springer-Verlag, 1989, vi + 562 pp, $69 (P). [ISBN: 0-387-51605-0] Citing 1960 papers by Kalman, Bellman, and Pontryagin as the beginnings of modern control theory and a shift in 1958 in the approach to filtering, the editors have assembled papers that commemorate what they take to be the entire history of system theory as the discipline is now defined and practiced. AWR

Systems Theory, T(18), P. *Nonlinear Control Systems: An Introduction, Second Edition.* Alberto Isidori. Communic. & Control Engin. Ser. Springer-Verlag, 1989, xii + 479 pp, $49.50. [ISBN: 0-387-50601-2] First appeared in 1983; is here rearranged on the basis of classroom use, supplemented with two new chapters, and given more of a differential geometry orientation. Supposes familiarity with linear system theory; intended as a graduate text and reference for professionals in the field. AWR

Probability, T(17: 2), P. *Topics in Advanced Econometrics: Probability Foundations.* Phoebus J. Dhrymes. Springer-Verlag, 1989, xii + 379 pp, $39.95. [ISBN: 0-387-97178-5] Contains probability theory that is needed for rigorous study of nonlinear econometrics. Measure theory, LLN and LLT for independent sequences and Martingales, ergodic theory. TH

Stochastic Processes, T(17-18). *Stochastic Differential Equations: An Introduction with Applications, Second Edition.* Bernt Øksendal. Universitext. Springer-Verlag, 1989, xv + 186 pp, $29.95. [ISBN: 0-387-51740-5] Changes from *First Edition* (TR, April 1986) include a reorganization of material and changes in presentation. MLR

Statistics, P. *Latent Variable Path Modeling with Partial Least Squares.* Jan-Bernd Lohmöller. Physica-Verlag (US Distr: Springer-Verlag), 1989, 283 pp, $58.40. [ISBN: 0-387-91363-7] Theoretical exposition, with algorithms and applications, of Herman Wold's "soft modelling" method of constructing linear-relation models, where "soft" refers to the distributional assumptions. Extensive bibliography. RSK

Statistics, P. *Lecture Notes in Statistics-58: Decomposition and Invariance of Measures, and Statistical Transformation Models.* Ole E. Barndorff-Nielsen, Preben Blaesild, Poul Svante Eriksen. Springer-Verlag, 1989, 147 pp, $19.50 (P). [ISBN: 0-387-97131-9] Theoretical monograph presuming some background in topology, group theory and differential geometry, in addition to parametric statistics. Presents the relevant mathematics of decomposition or disintegration of measures and construction of invariant measures and uses these results to present the key aspects of transformation models. RSK

Statistics, P. *Lecture Notes in Statistics-60: Nonparametric Curve Estimation from Time Series.* László Györfi, *et al.* Springer-Verlag, 1989, viii + 153 pp, $19.50 (P). [ISBN: 0-387-97174-2] Theoretical treatment of nonparametric estimation methods

(primarily kernel estimates) for nonlinear time series and regression analysis for density, distribution function, and hazard rate estimation with time-series (not i.i.d.) data. TH

Statistics, P. *Experimental Design in Biotechnology.* Perry D. Haaland. Stat.: Textbooks & Monographs, V. 105. Marcel Dekker, 1989, xv + 259 pp, $89.75. [ISBN: 0-8247-7881-2] Uses real data to illustrate how experimental design can be applied effectively to biotechnological problems. Primarily concerned with small screening experiments (using fractional factorial designs) and response surface experiments for optimization. Appendix contains a tabulation of useful designs (35 screening and 9 response surface). Note price. RSK

Statistics, P. *Lecture Notes in Statistics-54: Theory of Optimal Designs.* Kirti R. Shah, Bikas K. Sinha. Springer-Verlag, 1989, viii + 171 pp, $19.50 (P). [ISBN: 0-387-96991-8] Theoretical monograph presenting state-of-the-art results and recent advances in discrete optimal designs. RSK

Statistics, T(16-17: 1, 2), P*, L. *Statistical Methodology in the Pharmaceutical Sciences.* Ed: Donald A. Berry. Stat.: Textbooks & Monographs, V. 104. Marcel Dekker, 1990, viii + 578 pp, $125. [ISBN:0-8247-8117-1] Collection of 16 articles presenting a variety of statistical methods useful in pharmaceutical research. Each starts from first principles and progresses to state-of-the-art techniques, most illustrated with real data. Presumes an undergraduate course in statistical theory and methods. RSK

Computer Graphics, P, L. *Computational Geometry: Curve and Surface Modeling.* Su Bu-qing, Liu Ding-yuan. Transl: Chang Geng-zhe. Academic Pr, 1989, x + 295 pp, $39.95. [ISBN: 0-12-675610-4] Translation of the 1979 Chinese version. Concerned with computer-based representation analysis and synthesis of two- and three-dimensional shapes. Basic theories, geometric properties, and numerical algorithms of curves and surfaces widely used in computer-aided design and manufacture. Spline functions, parametric cubic spline curves; Bézier curves; spline surfaces; nonlinear splines; plane curves and "fairness." Readers need analytical geometry, calculus, and linear algebra. Approximately 170 papers published since 1979 have been added to an already lengthy reference list. JDEK

Artificial Intelligence, P. *Integration of Visual Modules: An Extension of the Marr Paradigm.* John (Yiannis) Aloimonos, David Shulman. Academic Pr, 1989, xvi + 322 pp, $39.95. [ISBN: 0-12-053020-1] An overview of the field of computer vision and what we can do with it from two researchers at the Computer Vision Laboratory and Computer Science Department at the University of Maryland. JAS

Computer Science, T(17-18: 1), P. *Topology via Logic.* Steven Vickers. Tracts in Theoret. Computer Sci., V. 5. Cambridge U Pr, 1989, 200 pp, $39.50. [ISBN: 0-521-36062-5] Ignore the points of a set, consider a family of subsets, and wallow in the parallelism between intersection/union and conjunction/disjunction to discover topology in logic. Such is the flavor of this foundation for certain aspects of formal language theory which is aimed at graduate students in computer science. A scattering of exercises and interesting bibliography. JAS

Computer Science, S. *Advances in Computers, Volume 29.* Ed: Marshall C. Yovits. Academic Pr, 1989, xi + 359 pp, $59.95. [ISBN: 0-12-012129-8] The twenty-ninth volume in a series begun in 1960 addressing long-range research issues in the field of computer science. This volume addresses five issues, four of which are technical and one of which is social. The technical issues are computer security, human-computer interaction, network protocols, and computer chess. The social issue is a study of the state of computing in the Soviet Union and how it compares with the United States. GMS

Computer Science, T. *Object-Oriented Systems Analysis: Modeling the World in Data.* Sally Shlaer, Stephen J. Mellor. Comput. Ser. Prentice-Hall, xii + 144 pp, (P). [ISBN: 0-13-629023-X] An informal presentation suitable for a practitioner-oriented workshop in information theory. JAS

Computer Science, T(16), P. *Object-Oriented Analysis.* Peter Coad, Edward Yourdon. Comput. Series. Prentice-Hall, 1990, xi + 232 pp, $29. [ISBN: 0-13-629122-8] The two most popular techniques for the specification and design of large software systems has been until recently functional decomposition and data flow analysis. Now a newer and significantly improved technique is being studied which appears to offer significant benefits. This method, called object-oriented systems analysis, decomposes the system into a collection of objects, functions on objects, and classes of objects. This decomposition, it is claimed, is more natural and easy to understand for users and programmers. The authors give a complete introduction to this new software design method. GMS

Computer Science, P, L. *Mathematics for Large Scale Computing.* Ed: J.C. Díaz. Lect. Notes in Pure & Appl. Math., V. 120. Marcel Dekker, 1989, xi + 345 pp, (P). [ISBN: 0-8247-8122-8] Twelve papers arising from a special session of the AMS meeting held in Denton, Texas in the fall of 1986. Covers specific research applications, development, and analysis of numerical algorithms and performance evaluation of algorithms on advanced architectures. The editors seek to give a representative picture of several aspects of large scale computing. SM

Computer Science, T(16-18: 1, 2). *Object-Oriented Software Construction.* Bertrand Meyer. Ser. in Comput. Sci. Prentice-Hall, 1988, xviii + 534 pp. [ISBN: 0-13-629049-3] A software engineering perspective using the language Eiffel (designed by the author with Simula for inspiration) as a notation for conveying the concepts of objected-oriented programming. A serious and detailed study of concepts, not a programming manual. Numerous open-ended discussion problems for the student. JAS

Computer Science, T(16-17: 1), L. *Introduction*

to Parallel Programming. Steven Brawer. Academic Pr, 1989, xi + 422 pp, $39.95. [ISBN: 0-12-128470-0] An informal, practical introduction to parallel programming techniques for use on shared-memory multiprocessors. Does not assume a particular hardware or software platform. Stresses concepts and fundamental techniques. AO

Computer Science, P. *Parallel Processing and Medium-Scale Multiprocessors.* Editor: Arthur Wouk. SIAM, 1989, 207 pp, $31.50 (P). [ISBN: 0-89871-238-6] Proceedings of a research workshop held at Stanford University in January 1986. The twelve papers included in this volume cover topics in systems programming, languages for parallel programming systems, and applications programming. AO

Applications, T(16-17: 1), L. *A Concrete Approach to Mathematical Modelling.* Michael Mesterton-Gibbons. Addison-Wesley, 1989, xx + 597 pp, $48.50. [ISBN: 0-201-12910-8] Aimed at students with no prior background in mathematical modelling. Deterministic and probabilistic models are given roughly equal treatment. Models are chosen from a wide variety of applications areas. Mathematical prerequisites include calculus, linear algebra, ordinary differential equations, and probability and statistics. AO

Applications (Biological Science), S(16), L. *Some Mathematical Questions in Biology: Models in Population Biology.* Ed: Allan Hastings. Lect. on Math. in the Life Sci., V. 20. AMS, 1989, vii + 123 pp, $28 (P). [ISBN: 0-8218-1170-3] In this work five papers appear which deal with different aspects of population biology: cultural evolution, multilocus population genetics, spatially structured population genetics, chaos and the dynamics of epidemics, and the dynamics of ecological systems. Common to all four works is the use of nonlinear difference equations and nonlinear differential equations. SM

Applications (Biological Science), P. *Applied Mathematical Ecology.* Eds: Simon A. Levin, Thomas G. Hallam, Louis J. Gross. Biomath., V. 18. Springer-Verlag, 1989, xiv + 491 pp, $59.50. [ISBN: 0-387-19465-7] Proceedings of the Second Autumn Course on Mathematical Ecology held November and December 1986 at the International Centre for Theoretical Physics in Trieste, Italy. Builds on the topics covered in an earlier publication *Mathematical Ecology,* edited by T.G. Hallam and S.A. Levin (TR, April 1987). Includes application of mathematical ecology in ecotoxicology, resource management, epidemiology, demography, and population ecology. RH

Applications (Economics), P. *Theoretical Market Areas under Euclidean Distance.* Pierre Hanjoul, Hubert Beguin, Jean-Claude Thill. Mono. Ser. Institute of Mathematical Geography (2790 Briarcliff, Ann Arbor, MI 48105), 1988, 162 pp, $15.95 (P). A research-level monograph describing a theory of market areas in the case of a duopoly competing inside the Euclidean plane endowed with Euclidean distance. AO

Applications (Engineering), P. *Symbolic Computation: Applications to Scientific Computing.* Ed: Robert Grossman. Frontiers in Appl. Math., V. 5. SIAM, 1989, x + 185 pp, $24.50 (P). [ISBN: 0-89871-239-4] The seven papers in this volume are from the NASA-Ames Workshop on the Use of Symbolic Methods to Solve Algebraic and Geometric Problems Arising in Engineering which took place January 15-16, 1987. These papers describe applications of computer algebra systems in the study of chaotic systems, fluid dynamics, nonlinear control systems, and robotics. AO

Applications (Information Theory), S*(14), L*. *Information Theory: Saving Bits.* William Sacco, *et al.* Contemp. Appl. Math. Janson, 1988, vi + 58 pp, $7.95 (P). [ISBN: 0-939765-25-X] A brief monograph in workbook format which explores efficient storage and transmission of information and error control coding. A very useful introduction to these two areas of information theory. CEC

Applications (Management), T(16: 1), S, C. *Managerial Decision Analysis.* Danny Samson. Richard D Irwin, 1988, xiv + 743 pp, $43.95. [ISBN: 0-256-06162-9] Focuses on the modeling and implementation issues of supporting managerial decisions with rigorous analysis. The treatment is not very mathematical, requiring only secondary school algebra. Includes the foundations of the subject along with chapters on sensitivity analysis, behavioral aspects of decision analysis, implementation issues, decision support systems, and complementary decision aids. Includes a decision analysis software package and documentation called ARBORIST that provides a tool for solving and formulating decision tree models. This package runs on a PC and takes some effort to understand how it works. The text does not depend on the software which serves as a supplementary tool. CEC

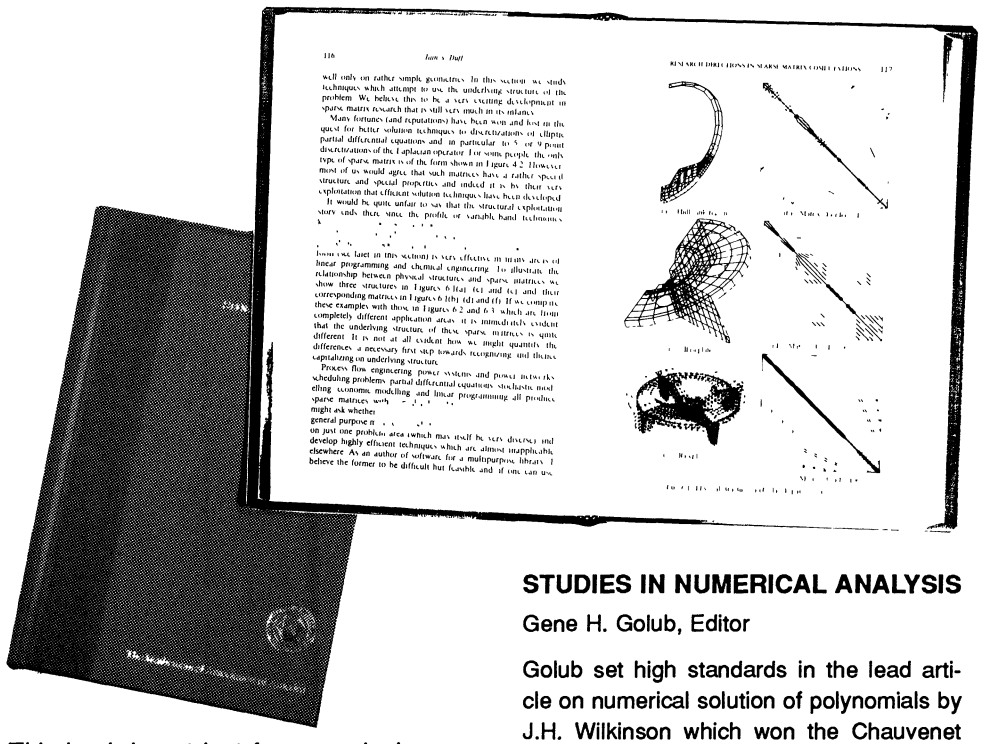
Applications (Physics), P. *Dynamics and Control of Multibody Systems.* Eds: J.E. Marsden, P.S. Krishnaprasad, J.C. Simo. Contemp. Math., V. 97. AMS, 1989, xvii + 468 pp, $46 (P). [ISBN: 0-8218-5104-7] Proceedings of the 1988 AMS-IMS-SIAM Joint Summer Research Conference. MU

Applications (Physics), P. *Lecture Notes in Physics-345: Schrödinger Operators.* Eds: H. Holden, A. Jensen. Springer-Verlag, 1989, 458 pp, $47.60. [ISBN: 0-387-51783-9] Fourteen papers on Schrödinger operators including short introductory lectures, longer survey articles, and current research topics. MPR

Reviewers

CEC: Clifton E. Corzatt, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; SM: Steve McKelvey, St. Olaf; AO: Arnold Ostebee, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; FR: Flauren Ricketts, Macalester; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; JAS: J. Arthur Seebach, Jr., St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.

STUDIES IN NUMERICAL ANALYSIS



This book is not just for numerical analysts, it is accessible to anyone with some mathematical foundation who wants to taste selections from a numerical analysis menu. The papers do a good job in motivating and introducing their subjects, and some provide a nice historical background. I recommend the articles in this book for anyone interested in learning more about the current research interests in numerical analysis. They may excite their readers to investigate further this relevant field of mathematics.

David C. Arney in *Mathematics and Computer Education*

STUDIES IN NUMERICAL ANALYSIS

Gene H. Golub, Editor

Golub set high standards in the lead article on numerical solution of polynomials by J.H. Wilkinson which won the Chauvenet Prize as the best piece of mathematical exposition to appear in 1985. These papers describe the wide range of research in numerical analysis. A variety of problems and of computational tools is treated. Some of the computations require only a hand-held calculator, while others require a powerful computer.

422 pp., 1985, ISBN 0-88385-126-1

List: $46.50 MAA Member: $34.50

Catalog Number MAS-24

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

A briefer version of Coughlin and Zitarelli's acclaimed Calculus with Applications—the most applied calculus text available

BRIEF CALCULUS WITH APPLICATIONS

Raymond F. Coughlin and David E. Zitarelli both of Temple University

Coughlin and Zitarelli's *Calculus with Applications* was an overnight success when it was introduced earlier this year. Instructors praised its fact-based, real-world case studies...its abundant applications exercises... its unusually thorough coverage.

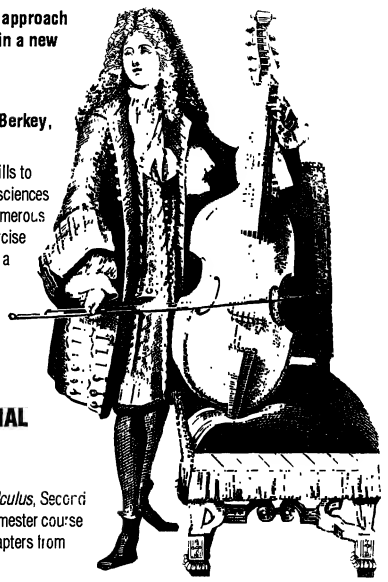
Now these same great features are available in a text for one-semester courses. *Brief Calculus with Applications* is the text professors have been asking for ever since the authors' longer book first appeared.

The mathematically honest approach to applied calculus is back in a new second edition...

APPLIED CALCULUS

Second Edition– Dennis D. Berkey, Boston University

Students need basic calculus skills to succeed in business, the social sciences or the life sciences. Through numerous worked examples, extensive exercise sets, abundant applications, and a uniquely clear, straightforward presentation, Berkey delivers these skills.



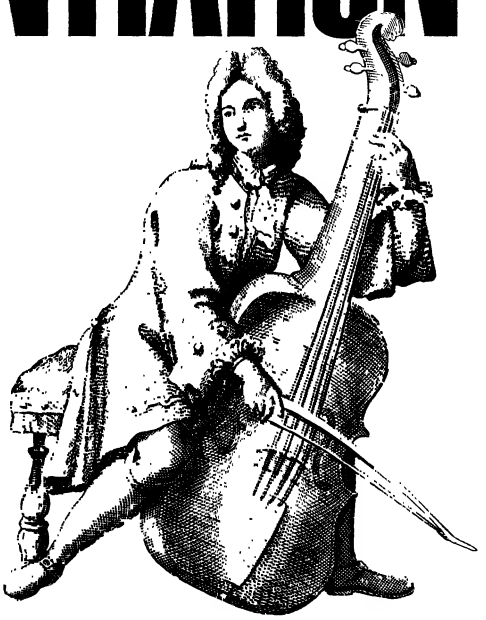
Also from Dennis D. Berkey

CALCULUS FOR MANAGEMENT, SOCIAL AND LIFE SCIENCES

Second Edition

This brief version of *Applied Calculus*, Second Edition, is perfect for the one-semester course. It includes seven appropriate chapters from *Applied Calculus*.

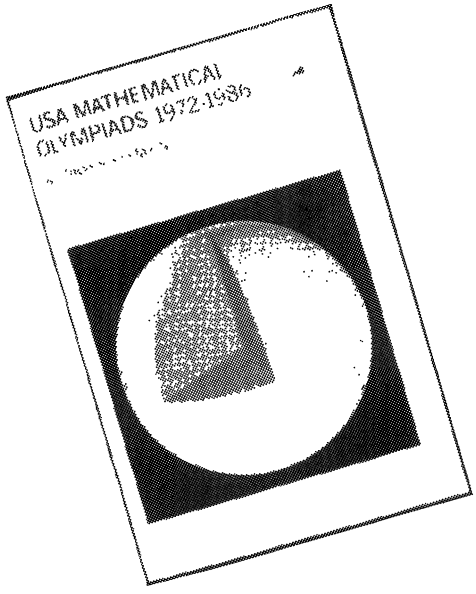
DIFFERENTIATION



SAUNDERS COLLEGE PUBLISHING

THE POWERS OF MATHEMATICS

USA MATHEMATICAL OLYMPIADS



Every year 100 of the most mathematically talented high school students in the country compete in the USA Mathematical Olympiad (USAMO). The USAMO is the third stage of a three-tiered mathematical competition for high school students in the United States and Canada that begins with the AHSME taken by over 400,000 students, continues with the American Invitational Mathematics Exam involving 2,000 students, and culminates in the 100-contestant USAMO.

USA MATHEMATICAL OLYMPIADS 1972-1986, PROBLEMS AND SOLUTIONS

Compiled by Murray S. Klamkin

People delight in working on problems "because they are there, " for the sheer pleasure of meeting a challenge. This is a book full of such delights. In it, Murray S. Klamkin brings together 75 original USA Mathematical Olympiad (USAMO) problems for years 1972-1986, with many improvements, extensions, finger exercises, open problems, references and solutions, often showing alternative approaches. The problems are coded by subject and solutions are arranged by subject as an aid to those interested in a particular field. Contains a glossary of frequently used terms and theorems, and a comprehensive bibliography with items numbered and referred to in brackets in the text. The problems are intriguing and the solutions elegant and informative. Students and teachers will enjoy working these challenging problems. Indeed all those those who are mathematically inclined will find many delights and pleasant challenges in this book.

180 pp., I988, ISBN-0-88385-634-4

List: $13.50 MAA Member: $12.50

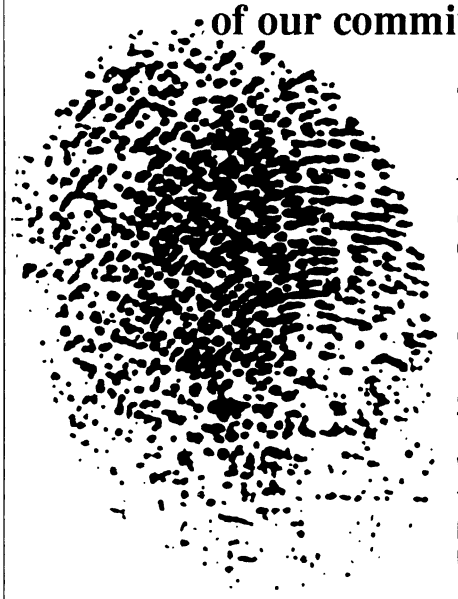
Catalog Number NML-33

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

Proof Positive

of our commitment to Mathematics



AN INTRODUCTION TO THE HISTORY OF MATHEMATICS, SIXTH EDITION
Howard Eves, University of Maine and University of Central Florida

This classic introduction to the history of mathematics is specially designed for undergraduate students. The completely revised sixth edition includes new information on women in mathematics, more illustrations and portraits, additional problem studies, and new "Cultural Connections" that show students how mathematics is part of our cultural heritage.

ABSTRACT ALGEBRA: AN INTRODUCTION
Thomas W. Hungerford, Cleveland State University

Written specifically for undergraduates, *Abstract Algebra* can be used from one to three terms. Presenting rings before groups, this text provides clear explanations, many examples, and abundant exercises, making this subject both interesting and accessible.

MATHEMATICAL LOGIC: APPLICATIONS AND THEORY
Jean E. Rubin, Purdue University

To make mathematical logic understandable for undergraduate students, this text teaches a specific proof writing system, introduces new topics with illustrative examples, incorporates a large number of exercises, and features abundant applications of logic to mathematics.

INTRODUCTION TO PROOF IN ABSTRACT MATHEMATICS
Andrew Wohlgemuth, University of Maine

In this original new text, inference rules are used as guides in setting up proofs. Students learn about proofs by actually doing them, following a logical sequence of definitions, rules, and theorems.

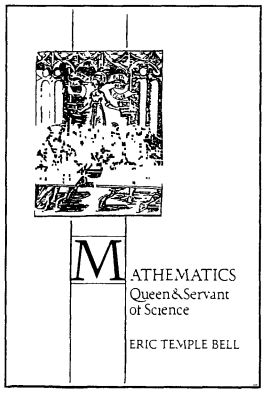


Saunders College Publishing

THE POWERS *OF* MATHEMATICS

The Curtis Center, Independence Square West, Philadelphia, PA 19106-3399

MATHEMATICS & BIOGRAPHY



MATHEMATICS: QUEEN AND SERVANT OF SCIENCE

E.T. Bell

An absorbing account of pure and applied mathematics from the geometry of Euclid to that of Riemann and its application in Einstein's theory of relativity. The twenty chapters treat such topics as: algebra, number theory, logic, probability, infinite sets and the foundations of mathematics, rings, matrices, transformations, groups, geometry, and topology. Republished in 1987 with corrections and an added Foreward by Martin Gardner.

454 pp., ISBN 0-88385-446-3
Paperbound

List: $15.95 MAA Member: $11.95

Catalog Number QAS

The book deserves a place in today's market. It is a much more popular work than most histories of the subject, and that is exactly what makes it accessible to students as well as to non-mathematicians. It is rewarding reading for . . . teachers and students at all mathematical levels.

Morris Kline of The Courant Institute

I WANT TO BE A MATHEMATICIAN
An Automathography in Three Parts

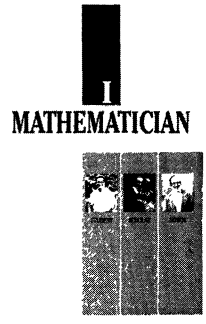
Paul R. Halmos

This is a book to be read with interest by all those who know, or might want to know, what mathematicians and mathematical careers are like. Paul R. Halmos begins with his school days and carries the reader swiftly through a career that has sustained itself at a high level since his first post-doctoral days at the Institute for Advanced Study in 1939, where he worked with John von Neumann among others. Still going strong in 1988, Halmos has contributed much to logic, operator theory, ergodic theory, and the literature in general.

442 pp., 1988, Paper, ISBN 0-88385-445-7

List: $18.00 MAA Member: $15.00

Catalog Number IWM



It is a truly unique book, which nobody but Paul Halmos could have written. I think it will be a classic.

Constance Reid

The book is exciting, witty, and well worth the time invested in its study. It communicates what it means to be a mathematician.

John Dossey in *The Mathematics Teacher*

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036
(202) 387-5200

Studies in Mathematical Economics

Volume 25 in the MAA Studies in Mathematics
Edited by Stanley Reiter

420 pp. Hardbound
ISBN-0-88385-027-X

List: $42.00
MAA Member: $31.00

"For the mathematician desiring to become familiar with modern mathematical, microeconomic theory, this volume is indispensable."
Robert Rosenthal
SUNY, Stony Brook
Department of Economics

Stanley Reiter, as editor, has brought together a distinguished group of contributors in this volume, in order to give mathematicians and their students a clear understanding of the issues, methods, and results of mathematical economics. The range of material is wide: game theory; optimization; effective computation of equilibria; analysis of conditions under which economies will move to the greatest possible efficiency under various forces, and the requirements for the flow of information needed to achieve efficient markets.

The material is interesting at all mathematical levels. For example, the initial article shows how even mathematically simple, concrete, two-person, nonzero sum games present us with the complexities and dilemmas of choices in real life. At the other extreme, the final article, by Debreu, begins by using the power of Kakutani's fixed point theorem to prove the existence of economic equilibria. In between, the reader will find beautiful uses of calculus, topology, combinatorial topology, and other topics.

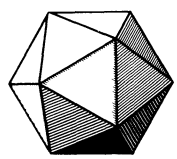
The chapters of this volume can be read independently, although they are related. The book begins with Meyerson's chapter on game theory and its theoretic foundations. The second chapter, by Simon, starts with the familiar criteria for maxima from calculus and goes on to develop more general tools of mathematical economics, including the Kuhn-Tucker and related conditions. The third contribution, by Mas-Collell, uses the tools of differential topology, including Sard's theorem, to study the competitive equilibria of whole families of economies using a differentiable point of view. Next Kuhn, building on the work of Scarf, shows how methods based on Sperner's lemma can be used to compute equilibria.

The next two chapters by Reiter and Hurwicz explore the properties of systems that are not purely competitive. They bring analytical and topological tools to bear to determine what conditions on the exchange of information are needed to allow such markets to become optimally efficient.

Radner addresses one consequence of what Herbert Simon calls "bounded rationality." Managers neither know all the facts nor do they have unlimited ability to calculate. How should they allocate their time? The tools used to answer this question are fittingly probabilistic.

In the final chapter, Debreu gives four examples of mathematical methods in economics. These four examples alone give a sense of the breadth and nature of the field.

In this study, Reiter and his other contributors show the reader the subtlety and complexity of the subject along with the precision and clarity that mathematics bring to it.



ORDER FROM
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036

TWO BOOKS ON MATHEMATICS EDUCATION

TOWARD A LEAN AND LIVELY CALCULUS
Report of the Conference/Workshop to Develop Curriculum and Teaching Methods for Calculus at the College Level

Ronald G. Douglas, Editor
MAA Notes #6
249 pp., Paperbound, 1987,
ISBN-0-88385-056-7
List: $12.50
Catalog Number NTE-06

Should calculus be taught differently? Can it? Common wisdom says "no"—which topics are taught, and when, are dictated by the logic of the subject and by client departments. The surprising answer from a four-day Sloan Foundation-sponsored conference on calculus instruction, chaired by Ronald Douglas, is that significant change is possible, desirable and necessary. Meeting at Tulane University in New Orleans in January, 1986, a diverse and sometimes contentious group of twenty-five faculty, university, and foundation administrators, and scientists from client departments put aside their differences to call for a leaner, livelier, more contemporary course, more sharply focused on calculus's central ideas and on its role as the language of science.

This volume contains the results of that conference and the papers presented to the conferees. These are certain to be the point of departure and basis for efforts to strengthen and reshape calculus in the next decade.

CALCULUS FOR A NEW CENTURY: A PUMP NOT A FILTER

L.A. Steen, Editor
MAA Notes #8
272 pp., Paperbound, 1987,
ISBN-0-88385-058-3
List: $12.50
Catalog Number NTE-08

Proceedings of the Colloquium held in October 1987 in Washington, D.C. to discuss calculus reform. The mathematical community is challenged to make the introductory calculus course into a pump that feeds more students into science and engineering, not a filter that cuts down the flow. These proceedings, with contributions from over eighty authors, show the full sweep of concerns and approaches of all the groups involved in calculus reform, including those currently teaching traditional and innovative courses, those whose students or employees need to use calculus as a tool, and the department chairs, deans, and others who must mobilize the resources needed for this reform.

THE MATHEMATICAL ASSOCIATION OF AMERICA

ORDER FORM

Please send the following books:

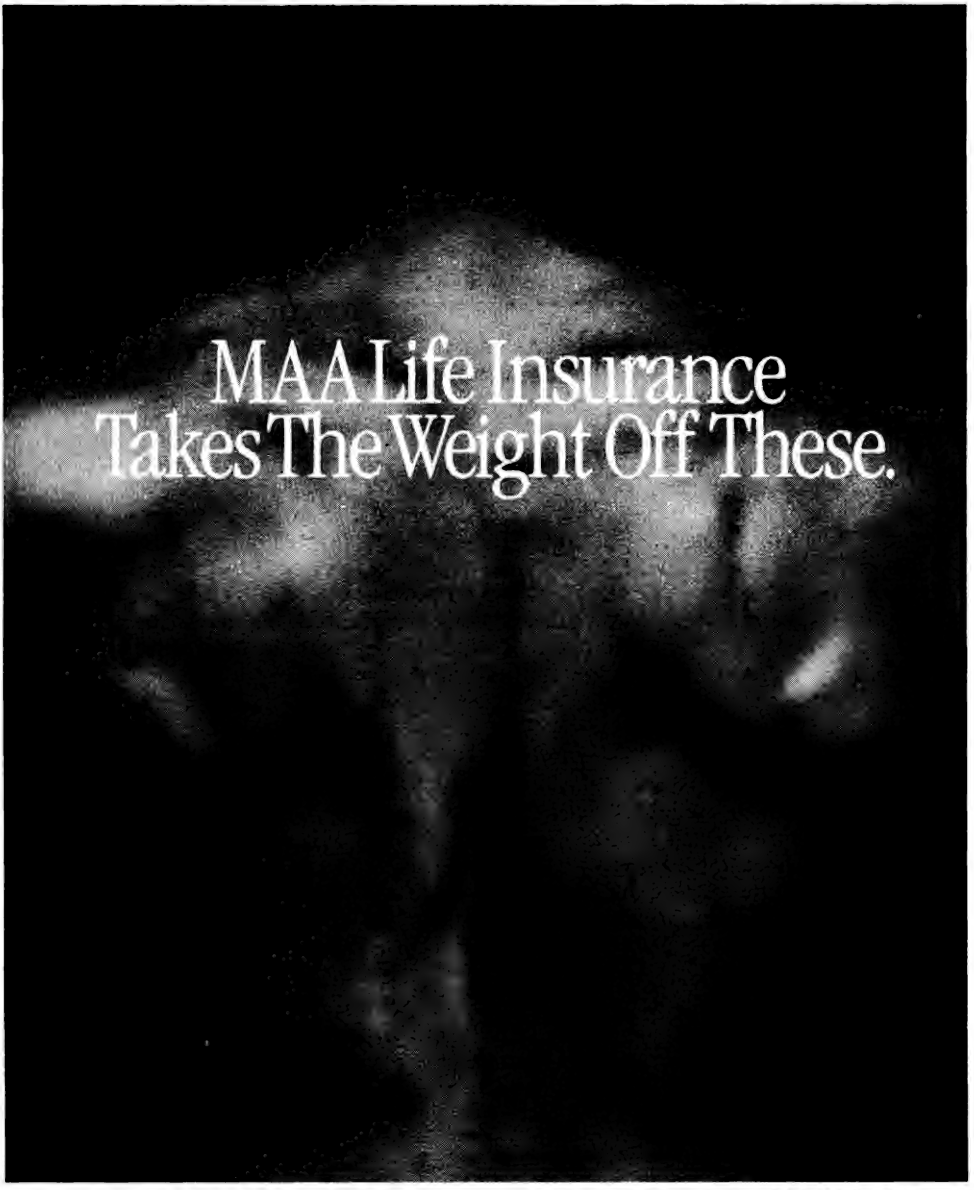
Quantity	Catalogue Number	Unit Price	Price
____	NTE-06	$12.50	_____
____	NTE-08	$12.50	_____
		Total $	_____

Name: _____

Address: _____

☐ Payment enclosed (sent postpaid)
☐ Please bill me (postage and handling extra)

Mail to: Mathematical Association of America, 1529 Eighteenth St., N.W., Washington, DC 20036



Your Association uses group purchasing power to negotiate top quality Term Life Insurance, at a very low price. What's more, it's insurance specially designed for your professional needs. It can be tailored for the individual, and it can stay with you even if you change jobs.

To take advantage of this benefit of membership—and take the weight off your shoulders—call 1-800-424-9883 for further details (in Washington, D.C. call 457-6820).

MAA INSURANCE
Designed by Members.
For Members.

The MAA Life Plan is underwritten by Connecticut General Life Insurance Company, A CIGNA Company, Hartford, Connecticut 06152 on form number GM 5000

Winner of the First MAA Book Prize . . .



Carus Mathematical Monograph, #19

Field Theory and its Classical Problems,

by Charles R. Hadlock

323 + xvi pp. Hardbound

List: $28.00 MAA Member: $22.00

Field Theory and its Classical Problems is one of those rare monographs that will capture and hold a mathematically prepared reader (one reasonably familiar with basic calculus and linear algebra) until the last page. Mathematicians will be intrigued by the development which places Galois Theory in an historical perspective and allows it to unfold from the emergence of the classical construction problems to the discovery of their solutions. The book presents some of history's greatest and most elegant mathematics.

Read what some reviewers say about this monograph:

"Hadlock has produced a pedagogic masterpiece . . . His idea of beginning with the three Greek problems and letting them fire the reader's spark plugs is brilliant . . . (his) ability to inject vitality and enthusiasm into mathematical text is remarkable. (George Piranian, University of Michigan)

"In the preface, the author claims to have written the book for himself . . . Fortunately for us, he chose to share his work with the mathematical community. I suggested the book as collateral reading in a one-semester course in Galois Theory and the students found it very readable and most helpful in establishing a motivation and direction . . . This is a delightful book for both student and teacher." (John D. Leadley, ***Mathematical Reviews)***

Table of Contents

Chapter 1—The Three Greek Problems
Constructible Lengths; Doubling the Cube; Trisecting the Angle; Squaring the Circle; Polynomials and Their Roots; Symmetric Functions; the Transcendence of π.

Chapter 2—Field Extensions
Arithmetic of Polynomials; Simple, Multiple, and Finite Extensions; Geometric Constructions Revisited; Roots of Complex Numbers; Constructibility of Regular Polygons I; Congruences; Constructibility of Regular Polygons II.

Chapter 3—Solution by Radicals
Statement of the Problem; Automorphisms and Groups; The Group of an Extension; Two Fundamental Theorems, Galois' Theorem; Abel's Theorem; Some Solvable Equations.

Chapter 4—Polynomials With Symmetric Groups
Background Information; Hilbert's Irreducibility Theorem; Existence of Polynomials over Q with Group S_n.



Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

f(z) Complex Variables

Complex plane or Riemann sphere.

See how functions combine - apply composition, integration, differentiation - in any combination.

***f(z)* can handle poles and knows how to pick a continuous branch of the logarithm.**

Generate animation as t paramter varies.

Watch the actual complex "graph" as it rotates in *four* dimensions!

Includes *f(z)*-Mandelbrodt

From our mail...

"Your f(z) software is one of the most interesting mathematical tools I have come across...It's a whole new mathematical world." S. S. M.

"Excellent, beyond my imaginings...many thanks once again for reopening my eyes to the many beauties of mathematics. . . I had never dreamed it would be so readily and easily available." D. H. - Australia

"f(z) is spectacular..I only wish I had had this when I was back in college!!!" V. W. - Raychem

"Great program, terrific graphics, never seen anything quite like it before." S. P.

For PC or Macintosh

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

Casson's Invariant for Oriented Homology 3-Spheres

An Exposition

Selman Akbulut and John D. McCarthy

In the spring of 1985, A. Casson announced an interesting invariant of homology 3-spheres via constructions on representation spaces. In the fall of that same year, Selman Akbulut and John McCarthy held a seminar on this invariant. These notes grew out of that seminar. The authors have tried to remain close to Casson's original outline and proceed by giving needed details, including an exposition of Newstead's results.

Mathematical Notes, 36 Paper: $19.50 ISBN 0-691-08563-3

Analytic Pseudodifferential Operators for the Heisenberg Group and Local Solvability

Daryl Geller

Many of the operators one meets in several complex variables, such as the famous Lewy operator, are not locally solvable. Nevertheless, such an operator L can be thoroughly studied if one can find a suitable *relative* parametrix—an operator K such that LK is essentially the orthogonal projection onto the range of L. The analysis is by far most decisive if one is able to work in the real analytic, as opposed to the smooth, setting. Here the author develops an analytic calculus for the Heisenberg group.

Mathematical Notes, 37 Paper: $29.50 ISBN 0-691-08564-1



Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900

ORDERS 800-PRS-ISBN (777-4726)



Carus Mathematical Monograph No. 20

THE GENERALIZED RIEMANN INTEGRAL

Robert M. McLeod
275 + xiii pages. Hardbound.
List: $24.95 MAA Member: $22.00

Until recently the most powerful and beautiful tools of integration theory have been accessible only to the privileged few whose studies extended through the Lebesgue integral. Now a new integral, a generalization of the familiar Riemann integral, has been discovered which has all the power and range of the Lebesgue integral but which can be readily understood by anyone who has studied calculus through the sophomore level.

The Generalized Riemann Integral is the first book-length presentation of this exciting new development in integration theory. Because of the clarity and organization of the exposition and the inclusion of exercises designed to actively engage the reader in the material, it is eminently suitable for use as a textbook at the advanced undergraduate or beginning graduate level. Furthermore, because it presents within a single volume results which were previously scattered throughout many research publications, it will undoubtedly also be of considerable interest to specialists in integration theory.

List of Contents: Definition of the Generalized Riemann Integral • Basic Properties of the Integral • Absolute Integrability and Convergence Theorems • Integration of Subsets of Intervals • Measurable Functions • Multiple and Iterated Integrals • Integrals of Stieltjes Type • Comparison of Integrals • Appendix: Solutions of In-Text Exercises

THE MATHEMATICAL ASSOCIATION OF AMERICA

Order From:
Mathematical Association of America
1529 Eighteenth Street NW
Washington DC 20036

NEW & CLASSIC TEXTBOOKS

Principles of Real Analysis
Second Edition

Charalambos D. Aliprantis and Owen Burkinshaw
Purdue University, Indianapolis, Indiana

"A well-written introduction to measure and integration with ample exercises and examples. ...The text emphasizes and uses the algebraic structure of the various function spaces considered and this certainly enhances the appeal."
—THE AMERICAN MATHEMATICAL MONTHLY

An ideal text for senior-undergraduate and first-year graduate courses, this book treats all topics rigorously, with illustrated examples and more than 450 exercises of varying degrees of difficulty.

January 1990, c. 304 pages, $39.95
ISBN: 0-12-050255-0

Problems in Real Analysis
A Workbook with Solutions

Charalambos D. Aliprantis and Owen Burkinshaw

Although based on the major textbook, **Principles of Real Analysis**, this book can be used as an independent source. It will be an invaluable tool for students who wish to develop a deep understanding and to acquire proficiency in the use of integration methods.

January 1990, c. 296 pages, $29.95
ISBN: 0-12-050256-9

Elements of Differentiable Dynamics and Bifurcation Theory

David Ruelle
Institute des Hautes Etudes Scientifiques
Bures-sur-Yvette, France

Written by an eminent researcher in the field, this introductory text presents the elements of the theory of differentiable dynamical systems to students of mathematics and the natural sciences in a lucid and yet rigorous way.

1989, 200 pages, $27.50
ISBN: 0-12-601710-7

Mathematics for Dynamic Modeling

Edward J. Beltrami
State University of New York at Stony Brook

"This is an excellent book aimed at undergraduates studying mathematical models which involve differential equations....The author has read a lot, and produced a very interesting variety of models to illustrate his basic themes of dynamical modelling. Within his self-imposed restriction to an emphasis on the dynamic aspects, the book is very successful, and provides a useful source for undergraduate courses."
—MATHEMATICAL REVIEWS

1987, 277 pages, $35.00
ISBN: 0-12-085555-0

Geometric Measure Theory
A Beginner's Guide

Frank Morgan
Williams College
Williamstown, Massachusetts

"Morgan's book is a pedagogical masterpiece."
—ELLIOTT LIEB

1988, 153 pages, $24.50
ISBN: 0-12-506855-7



ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
Book Marketing Department #S03030
1250 Sixth Avenue, San Diego, CA 92101

CALL TOLL FREE
1-800-321-5068
Quote this reference number for *free postage and handling* on your prepaid order ➥ **S03030**
or
Call 619-699-6400 for textbook adoption information.

Prices subject to change without notice. ©1990 by Academic Press, Inc.
All Rights Reserved. TS —S03030



Random Walks and Electric Networks,

by J. Laurie Snell and Peter Doyle

xiii + 159 pages. Hardbound

List: $26.50 MAA Member: $21.00

In this newest addition to the Carus Mathematical Monographs, the authors examine the relationship between elementary electric network theory and random walks, at a level which can be appreciated by the able college student. We are indebted to them for presenting this interplay between probability theory and physics in so readable and concise a fashion.

Central to the book is Polya's beautiful theorem that a random walker on an infinite street network in d-dimensional space is bound to return to the starting point when d = 2, but has a positive probability of escaping to infinity without returning to the starting point when d = 3. The authors interpret this theorem as a statement about electric networks, and then prove the theorem using techniques from classical electrical theory. The techniques referred to go back to Lord Rayleigh who introduced them in connection with an investigation of musical instruments.

In Part I the authors restrict themselves to the study of random walks on finite networks, establishing the connection between the electrical concepts of current and voltage and corresponding descriptive quantities of random walks regarded as finite state Markov chains. Part II deals with the idea of random walks on infinite networks.

Table of Contents

***Part I:* Random Walks on Finite Networks**

Random Walks in One Dimension
Random Walks in Two Dimensions
Random Walks on More General Networks
Rayleigh's Monotonicity Law

***Part II:* Random Walks on Infinite Networks**

Pólya's Recurrence Problem
Rayleigh's Short-Cut Method
The Classical Proofs of Polya's Theorem
Random Walks on More General Infinite Networks



Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

A D D I S O N - W E S L E Y

A broad spectrum of math texts for every level of learning!

Calculus like you've never seen it before!

Calculus
Ross L. Finney and George B.Thomas Jr., both of *Massachusetts Institute of Technology*

Calculus with Analytic Geometry, Third Edition
John B. Fraleigh, *University of Rhode Island*

Calculus and Analytic Geometry, Seventh Edition
George B. Thomas Jr. and Ross L. Finney, both of *Massachusetts Institute of Technology* (1988)

We're leading the way in precalculus!

Algebra and Trigonometry, Fifth Edition
M. L. Keedy and M. L. Bittinger, *Purdue University*, and *Indiana University-Purdue University at Indianapolis*, respectively.

Precalculus: Functions and Graphs
Franklin Demana and Bert K. Waits, both of *The Ohio State University*

Precalculus: Functions and Graphs
Jimmie Gilbert and Linda Gilbert, both of *University of South Carolina-Spartanburg*

Also on our math list for 1990. . .

Elementary Algebra: Concepts and Applications, Third Edition
M. L. Bittinger and M. L. Keedy; *Indiana University-Purdue University at Indianapolis*, and *Purdue University*, respectively.

Intermediate Algebra: Concepts and Applications, Third Edition
M. L. Bittinger, M. L. Keedy, and David Ellenbogen; *Indiana University-Purdue University at Indianapolis*, *Purdue University*, and *St. Michael's College*, respectively.

Linear Algebra, Second Edition
John B. Fraleigh, *University of Rhode Island*

Foundations of Higher Mathematics: Exploration and Proof
Daniel Fendel with Diane Resek, both of *San Francisco State University*

Applied Combinatorics with Problem Solving
Bradley W. Jackson and Dimitri Thoro, both of *San Jose State University*

Elementary Geometry from an Advanced Standpoint, Third Edition
Edwin Moise, *Emeritus, Queens College of the City University of New York*

Vector Analysis for Engineers and Scientists
P. E. Lewis and J. P. Ward, both of *Loughborough University of Technology (U.K.)*



Addison-Wesley Publishing Company
1 Jacob Way • Reading, MA 01867 • (617) 944-3700

From the publisher that's serious about math.

MAA STUDIES IN MATHEMATICS

Studies in Computer Science

Edited by Seymour Pollack
408 pp. Hardbound List: $33.50 MAA Member: $25.00

STUDIES IN COMPUTER SCIENCE continues the tradition of excellence of the respected series MAA Studies in Mathematics. Written by computer scientists for mathematicians, the book presents a readable and balanced discussion of the role of mathematics in computer science, and the contributions of computing to mathematics.

Opening with an historical overview of the development of computer science, the narrative continues with three articles on the nature of computer programs and the programming process, a discussion of computational complexity, and a concluding group of articles on the impact of computer science on artifical intelligence, numerical analysis, statistics, and simulation.

In the introduction the editor expresses his hope that the book will provide *"an interesting look at an explosive field in the process of becoming."* What an understatement! The process is well advanced; the impact of computer science on mathematics and its teaching is already immense. And this book is not just interesting, it is exciting and important as background reading for anyone interested in the interactions between mathematics and computer science. Teachers will find that it enriches their teaching and students will find it rewarding supplementary reading. Order your copy today!

Order From:
THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Pick a number, any number, and you pick excellence.

1 **Precalculus: A Problems-Oriented Approach, 3rd Edition** by David Cohen (University of California/Los Angeles) • Supplements: Instructor's Solutions Manual, Student Solutions Manual, Transparency Masters, Graph Toolz Software, Natural Language Mathematics Software by Mathens, Inc., Test Bank, West Math Test II

2 **Precalculus: With Unit Circle Trigonometry** by David Cohen (University of California/Los Angeles) • Supplements: Instructor's Solutions Manual, Student Solutions Manual, Transparency Masters, Graph Toolz Software, Natural Language Mathematics Software by Mathens, Inc., Test Bank, West Math Test II

3 **College Algebra, 3rd Edition** by Ralph C. Steinlage (University of Dayton) • Supplements: Instructor's Manual, Student Problem Solving Guide, Transparency Masters, Natural Language Mathematics Software by Mathens, Inc., Computerized Testing, Test Bank

4 **Understanding Elementary Algebra: A Course for College Students, 2nd Edition** by Arthur Goodman (Queens College of the CUNY) and Lewis R. Hirsch (Rutgers University) • Supplements: Instructor's Manual, Study Guide, Geometry Supplement to Accompany Understanding Elementary Algebra, Tutorial Software, Natural Language Mathematics Software by Mathens, Inc., Instructional Videotapes, Westest

5 **Understanding Intermediate Algebra: A Course for College Students, 2nd Edition** by Lewis R. Hirsch (Rutgers University) and Arthur Goodman (Queens College of the CUNY) • Supplements: Instructor's Manual, Study Guide, Geometry Supplement to Accompany Understanding Intermediate Algebra, Supplement on Sequences and Series, Tutorial Software, Natural Language Mathematics Software by Mathens, Inc., Instructional Videotapes, Westest

6 **Introductory Statistics** by Jay Devore and Roxy Peck (both from California Polytechnic State University-San Luis Obispo) • Supplements: Instructor's Manual, Student Solutions Manual, Transparency Masters, Test Bank, Westest

We're pleased to announce our 1990 math texts.

For more information on all our mathematics texts, write to West Publishing Company • College Department • 50 West Kellogg Boulevard • P.O. Box 64526 • St. Paul, Minnesota 55164-1003.

West Publishing Company

THREE IMPORTANT BOOKS ON MATHEMATICS EDUCATION

RESPONSES TO THE CHALLENGE: KEYS TO IMPROVED INSTRUCTION BY TEACHING ASSISTANTS AND PART-TIME INSTRUCTORS

The Committee on Teaching Assistants and Part-Time Instructors,
Bettye Anne Case, Chair

The committee that prepared this volume has been gathering information on policies, practices, successes, failures, and goals connected with the use of teaching assistants and part-time instructors. In this volume the committee presents and analyzes data showing who these teachers are, the extent and nature of their teaching duties, and the efforts made to assimilate them into the faculties. This volume will help you to see how your department compares nationally, to decide what steps you and your school should take, and to understand what additional resources might be needed.

280 pp., 1988, ISBN-0-88385-061-3
List: $15.00
Catalog Number NTE-11

GUIDELINES FOR THE CONTINUING MATHEMATICAL EDUCATION OF TEACHERS

Committee on the Mathematical Education of Teachers

These guidelines will be very useful to school teachers and supervisors, to college administrators who plan continuing education programs, to the college teachers who design and teach courses for teachers, and to school administrators who must think about requirements for continuing education of teachers. The guidelines are rich in specifics on course content, giving clear objectives for all courses. Teachers who want to dig out material for themselves or in order to enrich their classes will find the more than 500 references provided here under various topics an invaluable aid.

90 pp., 1988, ISBN-0-88385-060-5
List: $8.00
Catalog Number NTE-10

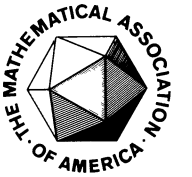
THE USE OF CALCULATORS IN THE STANDARDIZED TESTING OF MATHEMATICS

John W. Kenelly, Editor

The calculator is a universal tool for all those involved in quantitative work from science and engineering to business. Routine use of calculators is part of the training and testing of students headed for these fields. But this is not yet the case in mathematics. This symposium, jointly sponsored by the MAA and The College Board, sets out clearly the theoretical and practical issues that must be addressed as calculators are brought more fully into the mathematics curriculum. This is a practical group concerned with specific tests and test items. General theoretical considerations are set off by the specifics of individual test items and students' success rates on them. The Ohio Early College Mathematics Placement Test is reported on in detail by Joan R. Leitzel and Bert K. Waits. James W. Wilson and Jeremy Kilpatrick examine the theoretical issues in the development of calculator-based tests. John Harvey, now Chair of the MAA's Committee on Placement Examinations looks at the issues surrounding calculator use on placement examinations, as well as giving an overview of the symposium and a survey of developments through 1988.

vi + 50 pp., 1989, Copublished by the MAA and The College Board. LC No. 88-064-100

List: $6.50 Member: $6.50



Order from:

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 4 April 1990

Contents

(ISSN 0002-9890)

ARTICLES

Greed Is (Sometimes) Not Enough ANATOLE BECK 289

New Topologies from Old via Ideals . . . DRAGAN JANKOVIĆ AND T. R. HAMLETT 295

LETTERS TO THE EDITOR 311

UNSOLVED PROBLEMS

A Guessing Game of Bill Sands, and Bernardo Recamán's Barranca RICHARD K. GUY 314

NOTES

Polynomials with Large Numbers of Prime Values BETTY GARRISON 316

On Inequivalent Group Extensions F. B. CANNONITO 317

A Rearrangement Inequality and the Permutahedron A. VINCE 319

Musings on the Prime Divisors of Arithmetic Sequences PATRICK MORTON 323

THE TEACHING OF MATHEMATICS

Finding the Finite Groups of Symmetries of the Sphere MARJORIE SENECHAL 329

A Note on Defining Conditional Probability ARTHUR RAMER 336

Continuous Functions and Connected Graphs C. E. BURGESS 337

The l_2 Matrix Norm and the Spectral Radius ROBERT A. LESLIE 339

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 342

Advanced Problems and Solutions 350

REVIEWS

The Turing Omnibus: 61 Excursions in Computer Science. By A. K. Dewdney MARK BRIDGER 355

A Course in Constructive Algebra. By R. Mines, F. Richman, and W. Ruitenberg MICHAEL J. BEESON 357

TELEGRAPHIC REVIEWS 363

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

Greed Is (Sometimes) Not Enough

ANATOLE BECK,* *University of Wisconsin, Madison, WI 53706*

ANATOLE BECK: I was born in the Bronx and educated at Stuyvesant High School, Brooklyn College, and Yale University, where I studied with Shizuo Kakutani. After two postdoctoral years at Tulane and in Europe, I joined the faculty of the University of Wisconsin in 1958. I left Wisconsin in 1973 with the intention of spending my remaining years as Professor of Mathematics at the London School of Economics, returning after two years for personal reasons. I have retained my professional and personal ties to the LSE, however, and work there whenever my schedule permits.



Introduction. Many analyses in economic theory or in real life attempt to achieve global optimization by piecewise response to local considerations. Examples include the regulation of whole economies through the self-seeking strategies of individual participants. In the real world such theories are the subjects of heated debates [**3, 6**], not to say global thermonuclear confrontation [**4, 7**]. In the world of mathematics, we sometimes find examples in which the (plausible) argument that local advantage produces global optimization is not sustained by the facts.

We will deal with some geometric situations in which we seek to optimize a gain by a sequence of steps. In some cases, the choice of the best move at each step does not yield a global optimum. A simple example concerns the removal of non-overlapping squares from a rectangle so as to maximize the total area removed.

DEFINITION. *In the removal of areas from a geometric figure, the greedy strategy is the one which prescribes at each step the removal of the largest allowable area.*

DEFINITION. *The square of side s, in whatever position, will be denoted as* Q_s.

Example. In removing five squares from a 6×7 rectangle, the greedy strategy gives $1 \times Q_6, 4 \times Q_1$, for a total area of 40. By contrast, a combination of $1 \times Q_4, 2 \times Q_3, 2 \times Q_2$ gives a total of 42, the maximum possible.

DEFINITION. *A figure S is called gross if for every* $n \in \mathbb{N}$, *the removal of n figures of a specified sort is always best achieved by the greedy strategy. If, for any* $n \in \mathbb{N}$, *some other choice is better, then S is called subtle. In that case, the number n is called a subtlizer for S.*

A natural question, which has been circulated by word of mouth for a long time, is whether a square is gross or subtle with respect to the removal of discs [**1**]. We will not attempt to answer that question here. However, in the dual problem, there are infinitely many subtlizers for a disc with respect to the removal of squares.

The 6×7 rectangle is not unique for subtlety. In the first place, every rectangle close enough to it has the same property. More significantly, almost every rectangle has infinitely many subtlizers, as we shall see below.

*The research of this paper was supported in part by the Wisconsin Alumni Research Foundation and in part by the London School of Economics.

Triangles. We will begin with the problem of removing squares from an isosceles right triangle. Depending on the number of squares removed, it might pay to invest in a special arrangement for the first few squares. Consider FIGURE 1, where we see $3 \times Q_{1/3}$, $3 \times Q_{1/6}$, and $6 \times Q_{1/12}$, for a total of 12 squares. This removes all the triangle of side 1 except for 12 triangles of side $1/12$ with a total area of $12(1/2)(1/12)^2 = 1/24$. For the same triangle, the greedy strategy gives $1 \times Q_{1/2}$, $2 \times Q_{1/4}$, $4 \times Q_{1/8}$, and $5 \times Q_{1/16}$. This leaves 3 triangles of side $1/8$ and 10 of side $1/16$, for a total of $3(1/2)(1/8)^2 + 10(1/2)(1/16)^2 = 11/2^8$, which exceeds $1/24$. Thus, 12 is a subtlizer for the isosceles right triangle. Note that we have not proven that the alternative strategy is the best, but only that it is more effective than greed for $n = 12$.

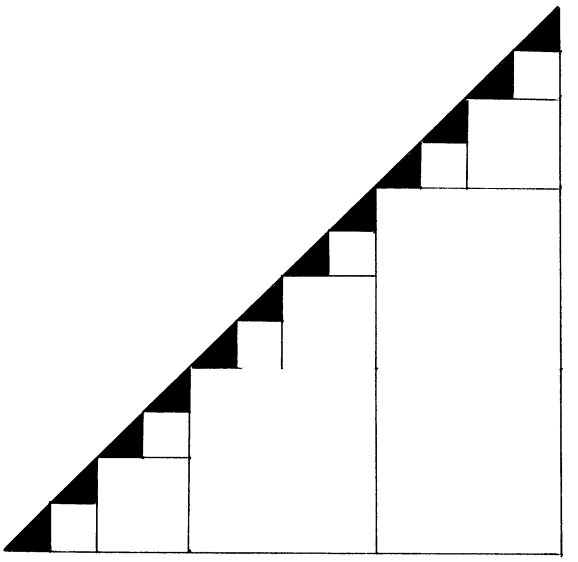


FIG. 1

We consider next the right triangle whose legs are 1 and $1 + \varepsilon$, and whose area is $A = (1 + \varepsilon)/2$. The largest square it contains has side $a = 1/2 + \varepsilon/(2\varepsilon + 4)$, and the removal of Q_a leaves two similar triangles whose hypotenuses are in the ratio a to $b = 1/2 - \varepsilon/(2\varepsilon + 4)$. Their areas are a^2A and b^2A respectively. The greedy strategy now chooses the largest square, which is in the first of these, and is Q_{a^2}. We will assume that ε is small enough so that $a^3 < b^2$, which will have as corollary $a^2 < b$. Then the third greedy square will be Q_{ab} from the smaller triangle. The next four greedy squares will be $1 \times Q_{a^3}$, $2 \times Q_{a^2b}$, and $1 \times Q_{ab^2}$, and the next 5 (since $a^4 < ab^2$) will be $1 \times Q_{a^4}$, $3 \times Q_{a^3b}$, and $1 \times Q_{a^2b^2}$. The removal of these twelve squares will leave triangles of the following sizes: $2 \times ab^2A$, $1 \times b^3A$, $1 \times a^4A$, $4 \times a^3bA$, $4 \times a^2b^2A$, and $1 \times ab^3A$. Thus, the total left by the greedy strategy will be the sum of those terms, which converges to $11/2^8$ as $\varepsilon \to 0$. On the other hand, the same squares chosen before by the alternate strategy (all lying within the isosceles right triangle) will leave over $1/24 + \varepsilon/2$, which converges to $1/24$. Thus,

for ε small enough the right triangle with legs 1 and $1 + \varepsilon$ is subtle, with subtlizer 12.

The number 12 is not unique. A similar phenomenon occurs at 24, 48, etc. In addition, the greedy strategy gains another $Q_{1/32}$ for $n = 25$, while the alternate gains only $Q_{1/48}$, but it is not enough to close the gap, nor is it for $n = 26$, but at $n = 27$, they are equally good. Similar analyses hold for $n = 49$, etc.

Actually, $3 \cdot 2^n + k$ is a subtlizer as long as

$$1/(3 \cdot 2^{n+1}) - k/(3 \cdot 2^{n+1})^2 < 1/2^{n+3} + 1/2^{n+4} - (k+1)/2^{2n+4},$$

i.e., $0 \leqslant k < (3 \cdot 2^n - 9)/5$. A similar calculation shows that $3 \cdot 2^n - k$ is a subtlizer so long as

$$1/(3 \cdot 2^{n+1}) + k/(3 \cdot 2^n)^2 < 1/2^{n+3} + 1/2^{n+4} + (k+1)/2^{2n+4},$$

i.e., $0 \leqslant k < (2^n + 3)/19$. Thus, all the numbers between $(56 \cdot 2^n - 3)/19$ and $9(2^{n+1} - 1)/5$ are subtlizers, $\forall n \in \mathbb{N}$. Dividing by $3 \cdot 2^{n+1}$, we see that the average density of subtlizers among the first N numbers exceeds $1/10$ for N large enough.

The role of the number 3 in this alternative strategy can be assigned to 5 as well. We can start with seven squares, $1 \times Q_{2/5}$ and $6 \times Q_{1/5}$, to leave five triangles of leg $1/5$. Then $5 \times Q_{10}$, $10 \times Q_{1/20}$, $20 \times Q_{1/40}$ and $40 \times Q_{1/180}$, leaving 80 triangles of side $1/80$, for a total area of $1/160$. Pursuing greed for 82 squares leaves 21 triangles of side 2^{-5} and 22 of side 2^{-6}. These have a total area of $53/2^{12}$, which is more than $1/160$. We can construct examples based on all odd numbers.

Our next example concerns a convex figure generated from the triangle with legs 1 and $1 + \varepsilon$ by replacing the hypotenuse by an arc of a circle. If the curvature is sufficiently small, the first n squares given by the greedy strategy are arbitrarily close to those in that triangle, as to both size and location. (The smallness of the curvature required depends on n and the degree of approximation desired.) Thus, if ε is small enough to make 12 a subtlizer for the triangle, then it will serve for the figure as well, when the curvature is very small. As an alternative to making the curvature small, we could take the legs small, but in the same ratio. Keeping the same value for ε (one which makes 12 a subtlizer for the triangle), we can choose a small value of $\delta > 0$ and draw a chord of length $\delta\sqrt{2 + 2\varepsilon + \varepsilon^2}$ in a circle of radius 1. Completing the triangle with legs of length δ and $\delta + \delta\varepsilon$, we get a figure similar to one of the previous ones, in which 12 is a subtlizer if δ is small enough.

Discs. We consider next the removal of squares from a disc. The first square removed is inscribed in the disc and its sides determine the orientation of all the other squares. Following the greedy strategy, the next square is one of the four which have one side included in a side of the first square, with the opposite side a chord of the circle, and the third, fourth, and fifth are the other three. The removal of these five squares leaves in the disc twelve areas, four of which are segments of the circle, while eight are figures of the sort described in the previous section. The sixth square is chosen from one of these by being placed so that two sides are included in the straight edges of the figure, while one vertex lies in the curved side (there is only one way to do this). The next seven squares are chosen from the seven figures congruent to that one in the same way, so that now twenty figures remain, four of which are segments. In this way, the greedy strategy selects its squares.

As the remaining pieces become smaller, we are especially interested in those in a particular portion of the circumference. Consider the four vertices of the first square, and observe that for each of them there is an arc in which the tangent to the disc at any point of the arc makes an angle with one side of the square whose tangent m is so near to 1 that the right triangle with sides 1 and m is subtle with subtlizer twelve. As the greedy strategy continues, the figures left over are just such triangles, modified by replacing the hypotenuses with arcs of the original circle. When these are small enough, and in the part of the disc where the slope of the circle is close enough to 1, then that figure will be subtle, with subtlizer twelve.

If we now choose one of these figures, T, and find the first value of n for which the greedy strategy has taken twelve squares out of T, then a change of those twelve squares into twelve others will increase the area extracted, making n subtlizer for the disc. Clearly, this happens for infinitely many values of n.

It follows that the greedy strategy will eventually leave subtle figures as components along the indicated arcs. For any such figure, greed will eventually select twelve squares, and these will be the twelve which it would select if there were not other places to choose from. The fact that other squares in other parts of the disc will be chosen between these is irrelevant. However long it takes to choose the twelfth, that number is a subtlizer for the disc since just replacing those twelve (while keeping all the others) is already more efficient than greed.

Of course, this analysis gives no hint of which number n is the first subtlizer, and indeed, no actual subtlizer is known, but it is clear that there are infinitely many.

Rectangles. If discs are complicated (those unruly round edges!), then perhaps rectangles are simple. We have already seen, in the 6×7 rectangle, that they are not all gross. In general a rectangle $(n^2 + n) \times (n^2 + n + 1)$ can be exhausted by $2n + 1$ squares ($1 \times Q_{n^2}$, $n \times Q_n$, and $n \times Q_{n+1}$), but not if one is greedy (in which case one gets only $1 \times Q_{n^2+n}$ and $2n \times Q_1$). But these are very special cases. Note that the rectangle $6 \times (7 - \varepsilon)$ is also subtle providing that ε is small enough. Indeed, we can cover all but $6\varepsilon - \varepsilon^2$ by $1 \times Q_4$, $2 \times Q_2$, and $2 \times Q_{3-\varepsilon}$. The greedy strategy $(1 \times Q_6, 4 \times Q_{1-\varepsilon})$ covers all but $(2 + 4\varepsilon)(1 - \varepsilon) = 2 + 2\varepsilon - 4\varepsilon^2$, which is more unless $\varepsilon > (\sqrt{10} - 2)/3$, which is quite a bit. That is, every rectangle with a ratio of sides between $6/7$ and $d = 18/(19 + \sqrt{10})$ is subtle, and similar analyses apply to the other rectangles $(n^2 + n) \times (n^2 + n + 1)$ mentioned above.

On the other side, we can examine rectangles $(6 - \varepsilon) \times 7$, where the combination $1 \times Q_{4-5\varepsilon/4}, 2 \times Q_{2+\varepsilon/4}, 2 \times Q_{3-\varepsilon/2}$ covers all but $(4 - 5\varepsilon/4)(7\varepsilon/4) = 7\varepsilon - 35\varepsilon^2/16$. Greed gives $1 \times Q_{6-\varepsilon}$ and $4 \times Q_{1+\varepsilon}$, which leaves over $(1 + 2\varepsilon)(2 - 5\varepsilon) = 2 - 3\varepsilon - 10\varepsilon^2$, which is more unless $\varepsilon > 4(\sqrt{26} - 4)/25$, that is, for rectangles with ratio between $c = (6.64 - 4\sqrt{26}/25)/7$ and $6/7$. Thus, every rectangle with ratio between c and d is subtle.

However, there are other subtle rectangles. If $0 < \alpha < 1$ is the ratio of the sides of a rectangle, and $1 = n\alpha + \beta$, where $0 \leqslant \beta < \alpha$, then if the rectangle $\alpha \times \beta$ is subtle with subtlizer n_1, it follows that the original rectangle has $n + n_1$ as a subtlizer. Similarly if the rectangle $\alpha \times (\alpha + \beta)$ is subtle with subtlizer n_2, then $n + n_2 - 1$ is a subtlizer of the original rectangle. These observations allow us the following conclusions: We know that a rectangle R is subtle if its ratio x lies between c and d. Also, if $1/x = n_1 + r_1$, where $0 \leqslant r_1 < 1$, then R is subtle if $c < r_1 < d$. Recursively, $1/r_{j-1} = n_j + r_j$, where $0 \leqslant r_j < 1$, and R is subtle if

$c < r_j < d$ for any $j \in \mathbb{N}$. We now need to remark that for almost all x, $c < r_j < d$ for infinitely many $j \in \mathbb{N}$. This follows from the fact that the transformation $T(x) = 1/x - [1/x]$ is an ergodic transformation in $[0, 1)$ with respect to the measure μ generated by $\mu([0, t]) = \ln(1 + t)/\ln 2$ [3]. Thus, for almost all $0 < x < 1$, we have $c < r_j < d$ for a set of $j \in \mathbb{N}$ with asymptotic density $(\ln(1 + d) - \ln(1 + c))/\ln 2$. *A fortiori*, there are infinitely many subtlizers for each such x, since $n_1 + \cdots + n_j + 5$ is a subtlizer for each $j \in \mathbb{N}$ with $c < r_j < d$. Unfortunately, the density of numbers j for which the removal of j squares by greed leaves a rectangle of the indicated ratio is 0, since the expected value of n_j is infinite [2]. More to the point, I conjecture that the set of subtlizers of almost every rectangle has density 0, which would be true if it were known that for every rectangle $x \times 1$ with x small enough, and for every $n \in \mathbb{N}$, the optimal choice of squares includes at least one Q_x.

Discs and spheres. We return now to an old question. In removing discs from a square, is the square gross or subtle? To the best of my knowledge, the answer is not known. Even the fact that the greedy strategy will remove 100% of the area is an interesting problem, and has an elementary solution:

THEOREM. *If nonoverlapping open discs are removed from a square by the greedy strategy, the sum of the areas of the infinitely many discs removed will equal the area of the square.*

Proof. Among the discs removed by the greedy strategy are (first) the disc inscribed in the square and then the four discs in the four corners followed later (but not immediately) by the four in the corners, and (even later) by the four in the corners, ad infinitum. These clearly do not exhaust the area of the square, but leave infinitely many "trianguloids," bounded on two sides by tangent circles, and on the third by a circle or line tangent to the other two. The result we wish follows from the fact that if each such trianguloid has discs removed from it as shown in FIGURE 2, these will add up to more than half its area, leaving a new generation of

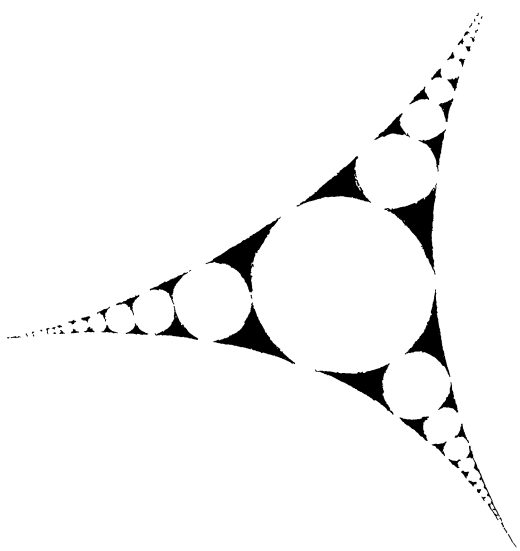


FIG. 2

trianguloids to be treated similarly. Since all the discs we are removing are included among those taken (eventually) by the greedy strategy, we see that infinitely many generations of this process leaves a set of area 0.

To show that the discs removed are more than half the area, we need only prove the fact for the "quadroid" marked in FIGURE 3. If all the sides of the original

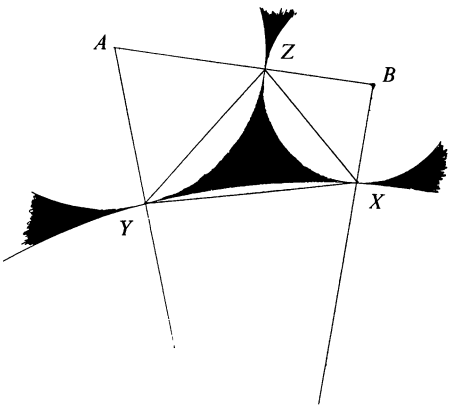


FIG. 3

trianguloid were arcs of circles, then the fact would follow from the lemma below, whose proof we leave as an amusement for the reader:

LEMMA. *If, in the triangle ABC, the inscribed circle meets the triangle in the points X, Y, and Z* ($AZ = AY$, $BZ = BX$, $CX = CY$), *then* $Area(XYZ) \leqslant (1/2)Area(AYXB)$.

If one side of the trianguloid is a line segment, then the same fact follows as a limit (or by an analogous argument). Thus we have a proof of this theorem by elementary geometry. QED

The method shown does not generalize to the removal of spheres from a cube, nor to higher dimensions. It is true, but not obvious, that the theorem holds in these dimensions, but the only proof I can construct is one which mimics the Vitali Covering Theorem (*mutatis mutandis*). A more elementary proof would be of interest.

Conclusion. In geometry, greed is sometimes not the best strategy. The lesson for other aspects of life is obscure.

REFERENCES

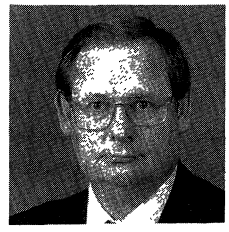
1. Paul Erdös and Donald J. Newman, Exhausting an area with discs, *Proceedings of the American Mathematical Society*, 45 (1974) 305–308.
2. A. Ya. Khinchine, Metrische Kettenbruchprobleme, *Compositio Mathematica*, 1 (1935) 361–382.
3. Robert Lekachman, Greed Is Not Enough: Reaganomics, Pantheon, New York, 1982.
4. Karl Marx, Das Kapital, Otto Meissner, Hamburg, 1867.
5. Alfred Rényi, Representations for real numbers and their ergodic properties, *Acta Mathematica Akademiae Scientiarum Hungaricae*, 8 (1957) 477–493.
6. Daniel Seligman, Good greed, *Fortune* (March 2, 1987) 119.
7. Adam Smith, An Inquiry Into The Nature And Causes Of The Wealth Of Nations, Strahan & Cadell, London, 1776.

New Topologies from Old via Ideals

DRAGAN JANKOVIĆ and T. R. HAMLETT, *East Central University, Ada, OK 74820*



DRAGAN JANKOVIĆ: I received my Ph.D. from the University of Belgrade, Yugoslavia, in 1983. Since then I have spent one year at the University of Arkansas, Fayetteville, and three years at Texas Tech University, Lubbock. At present I have a year's study leave from my position at East Central University, Ada, Oklahoma.



T. R. HAMLETT: I received my B.A. ('71) from Arkansas College and my M.S. ('73) and Ph.D. ('75) from the University of Arkansas under the direction of Paul Long. After three years military duty at Wright-Patterson AFB, Ohio, I spent three years at Arkansas Tech University, one year at Arkansas College, and am concluding my eighth year at East Central University, Ada, Oklahoma.

1. Introduction. The subject explored in this paper has a lengthy and interesting historical development along two lines which seem to converge in the 1975 paper of Samuels [**26**]. One line is a generalization, or unification, of the concept of closure point, ω-accumulation point (limit point in T_1 topological spaces), condensation point, and point of the second category. The other line is a study of the properties which spaces have "locally," and which, in several interesting cases, implies that the space has "globally." This last line is probably best exemplified by the 1945 paper of Vaidyanathaswamy [**31**].

Although the topic of ideals in general topological spaces is treated in the classic text by Kuratowski ([**16**], [**17**]) and also in [**30**], it is not found in most modern texts. Consequently, we have written this paper as a supplement to an advanced undergraduate, graduate, or topics course in General Topology, as well as a topic of interest to a specialist in General Topology. In addition, our subject has an excellent potential for application in other branches of mathematics. The Cantor-Bendixson Theorem (Theorem 5.2), Example 4.3, and Theorem 7.9 exemplify this potential.

In what follows, we will attempt to develop this topic in a logical, systematic fashion, and offer some new results, improvements of known results, and some applications. We have included examples liberally which are intended not only to reinforce the abstract concepts, but also to point out that ideals are present in some familiar places which the reader may have already visited.

We will use the following notational conventions: a "space" will always mean a topological space; given a space (X, τ) and a point $x \in X$, $\mathcal{N}(x)$ will denote the *open neighborhood system at* x (i.e., $\mathcal{N}(x) = \{U \in \tau : x \in U\}$); $\mathrm{cl}_\tau(A)$ (resp., $\mathrm{Int}_\tau(A)$) will denote the *closure* (resp., *interior*) *of* A *with respect to* τ (we will omit τ if there is no possibility of confusion). The following important ideals in a space (X, τ) will be denoted as follows:

$\mathcal{I}_f$—the ideal of finite subsets of X,
$\mathcal{I}_c$—the ideal of countable subsets of X,
$\mathcal{I}_{cd}$—the ideal of closed discrete sets in (X, τ),
$\mathcal{I}_n$—the ideal of nowhere dense sets in (X, τ),

$\mathscr{I}_m$—the ideal of meager sets in (X, τ),
$\mathscr{I}_s$—the ideal of scattered sets in (X, τ), and
$\mathscr{I}_k$—the ideal of relatively compact sets in (X, τ).

We will use the abbreviations "iff" for "if and only if" and "$\rightarrow$" to mean "implies." The symbol ■ is used to indicate the end (or omission) of a proof.

2. Fundamentals. Let us begin by defining an *ideal.*

Definition 2.1. A nonempty collection $\mathscr{I}$ of subsets of a set X is said to be an *ideal on* X, if it satisfies the following two conditions:
(A) $A \in \mathscr{I}$ and $B \subseteq A \rightarrow B \in \mathscr{I}$ (heredity).
(B) $A \in \mathscr{I}$ and $B \in \mathscr{I} \rightarrow A \cup B \in \mathscr{I}$ (finite additivity).

Note that the collection of complements of sets in a proper ideal (one not containing the space) will be a nonempty collection of nonempty sets closed under the operations of superset and finite intersection. Such a collection is called a *filter*; hence an ideal is sometimes called a *dual filter.*

It is a common practice to say that a space "has a property $\mathscr{P}$ locally at a point x" if x has some neighborhood U such that U has the property. Extending this idea to subsets, we say that a subset A *has the property* $\mathscr{P}$ *locally at a point* x if x has some neighborhood U such that $U \cap A$ has the property [**16**]. Note that the point x may or may not be in the subset A. Typically, ideals are defined as a collection of subsets with some particular property. The set of points in a space at which a subset does not have this property is a fundamental concept in what follows.

Definition 2.2 [**16**]. Let (X, τ) be a space with an ideal $\mathscr{I}$ on X. Then $A^\star(\mathscr{I}, \tau) = \{x \in X : A \cap U \notin \mathscr{I} \text{ for every } U \in \mathscr{N}(x)\}$ is called the *local function of* A *with respect to* $\mathscr{I}$ *and* τ.

When there is no ambiguity, we will write $A^\star(\mathscr{I})$ or simply $A^\star$ for $A^\star(\mathscr{I}, \tau)$ and call it the "local function of A." We remark that in Definition 2.2 $\mathscr{N}(x)$ may be replaced by the neighborhood system of x.

The simplest ideals are $\{\phi\}$ and $\mathscr{P}(X) = \{A : A \subseteq X\}$. Observe that $A^\star(\{\phi\}) = \mathrm{cl}(A)$ and $A^\star(\mathscr{P}(X)) = \phi$ for every $A \subseteq X$. Now let $\mathscr{I}_f$ be the ideal of finite subsets of X, and let A^ω be the set of *ω-accumulation* points of $A \subseteq X (x \in A^\omega$ iff $U \cap A$ is infinite for every $U \in \mathscr{N}(x))$. It is not difficult to check that $A^\omega = A^\star(\mathscr{I}_f)$. If $\mathscr{I}_c$ is the ideal of countable subsets of X, $A^\star(\mathscr{I}_c)$ is precisely the set of *condensation* points of $A (x \in X$ is a *condensation* point of $A \subseteq X$ if $U \cap A$ is uncountable for every $U \in \mathscr{N}(x))$. Thus $A^\star(\mathscr{I})$ is a generalization of closure point, ω-accumulation point and condensation point. We will consider these examples again later in this section.

The following theorem ([**16**], [**31**]), contains many basic and useful facts concerning the local function.

THEOREM 2.3. *Let* (X, τ) *be a space with* $\mathscr{I}$ *and* $\mathscr{J}$ *ideals on* X, *and let* A *and* B *be subsets of* X. *Then*
(a) $A \subseteq B \rightarrow A^\star \subseteq B^\star$,
(b) $\mathscr{I} \subseteq \mathscr{J} \rightarrow A^\star(\mathscr{J}) \subseteq A^\star(\mathscr{I})$,
(c) $A^\star = \mathrm{cl}(A^\star) \subseteq \mathrm{cl}(A)$ ($A^\star$ *is a closed subset of* $\mathrm{cl}(A)$),
(d) $(A^\star)^\star \subseteq A^\star$,
(e) $(A \cup B)^\star = A^\star \cup B^\star$,

(f) $A^\star - B^\star = (A - B)^\star - B^\star \subseteq (A - B)^\star$,
(g) $U \in \tau \to U \cap A^\star = U \cap (U \cap A)^\star \subseteq (U \cap A)^\star$, *and*
(h) $I \in \mathscr{I} \to (A \cup I)^\star = A^\star = (A - I)^\star$. ■

Let $\mathscr{P}(X)$ be the class of all subsets of X. If $(\)^c : \mathscr{P}(X) \to \mathscr{P}(X)$ is a function satisfying (1) $\phi^c = \phi$, (2) $A \in \mathscr{P}(X) \to A \subseteq A^c$, (3) $A \in \mathscr{P}(X)$, $B \in \mathscr{P}(X) \to (A \cup B)^c = A^c \cup B^c$, and (4) $A \in \mathscr{P}(X) \to (A^c)^c = A^c$; then $(\)^c$ is called a *Kuratowski closure operator* and $\{A \in \mathscr{P}(X) : A = A^c\}$ is a collection of closed sets for a topology on X.

It is not difficult to check that if d: $\mathscr{P}(X) \to \mathscr{P}(X)$ is a function satisfying
(1) $d(\phi) = \phi$,
(2) $d(A \cup B) = d(A) \cup d(B)$, and
(3) $d(d(A)) \subseteq d(A)$
then $(\)^c : \mathscr{P}(X) \to \mathscr{P}(X)$ defined by $A^c = A \cup d(A)$ is a Kuratowski closure operator on $\mathscr{P}(X)$. Note that d does not necessarily coincide with the derived set operator in the generated topology.

Since *: $\mathscr{P}(X) \to \mathscr{P}(X)$ satisfies all the required conditions for the function d (Theorem 2.3, (e), (d); $\phi^\star = \phi$), we have that $\mathrm{cl}^\star(A) = A \cup A^\star$ is a Kuratowski closure operator. It is not difficult, however, to verify directly that $\mathrm{cl}^\star$ is a closure operator. We will denote by $\tau^\star(\mathscr{I})$ the topology generated by $\mathrm{cl}^\star$ where τ is the original topology on X, that is, $\tau^\star(\mathscr{I}) = \{U \subseteq X : \mathrm{cl}^\star(X - U) = X - U\}$. When no ambiguity is present we will simply write $\tau^\star$.

We have already observed that if $\mathscr{I} = \{\phi\}$, then $A^\star = \mathrm{cl}(A)$. Hence in this case $\mathrm{cl}^\star(A) = \mathrm{cl}(A)$ and $\tau^\star = \tau$. If $\mathscr{I} = \mathscr{P}(X)$, then $A^\star = \phi$ for every $A \subseteq X$ and hence $\tau^\star(\mathscr{I})$ is the discrete topology. The ideals $\{\phi\}$ and $\mathscr{P}(X)$ illustrate extreme cases where $\tau^\star = \tau$ and $\tau^\star =$ discrete, respectively. Since for every ideal $\mathscr{I}$ on X we have $\{\phi\} \subseteq \mathscr{I} \subseteq \mathscr{P}(X)$, it follows from Theorem 2.3(b) that $\tau \subseteq \tau^\star \subseteq$ discrete. More specifically, if $\mathscr{I}$ and $\mathscr{J}$ are ideals on (X, τ) such that $\mathscr{I} \subseteq \mathscr{J}$, then $\tau^\star(\mathscr{I}) \subseteq \tau^\star(\mathscr{J})$ (apply Theorem 2.3(b)).

Our next example illustrates the possibility that $\mathscr{I} \neq \mathscr{P}(X)$ and yet $\tau^\star(\mathscr{I})$ is discrete.

Example 2.4. Let us denote by N the set of natural numbers. If we endow N with a "partition" topology τ generated by the base $\{\{2n - 1, 2n\} : n \in N\}$ and consider the ideal $\mathscr{I}_f$, then it is clear that $N^\star(\mathscr{I}_f) = \phi$. Therefore $\tau^\star(\mathscr{I}_f)$ is discrete. More generally, $X^\star(\mathscr{I}_f) = \phi$ iff (X, τ) is nearly discrete (i.e., each $x \in X$ has a finite neighborhood) iff $\tau^\star(\mathscr{I}_f)$ is discrete.

Let us agree to denote by A^d (resp., $A^{d^\star}$) the *derived* set of A in (X, τ) (resp., $(X, \tau^\star(\mathscr{I}))$ ($x \in A^d$, that is, x is a *limit* (or *accumulation*) point of A iff $(U - \{x\}) \cap A \neq \phi$ for every $U \in \mathscr{N}(x)$).

Obviously, $A^{d^\star} \subseteq A^d$ and also $x \in A^{d^\star}$ iff $x \in \mathrm{cl}^\star(A - \{x\})$ iff $x \in (A - \{x\}) \cup (A - \{x\})^\star$ iff $x \in (A - \{x\})^\star$. This means that $x \in A^{d^\star}$ iff $(U - \{x\}) \cap A \notin \mathscr{I}$ for every $U \in \mathscr{N}(x)$. Clearly, $A^{d^\star} \subseteq A^\star$. If $\{x\} \in \mathscr{I}$, then by Theorem 2.3 (h), $x \in A^{d^\star}$ iff $x \in A^\star$. The ideals in the following two examples have the property that $\{x\} \in \mathscr{I}$ for each $x \in X$.

Example 2.5. Let $\mathscr{I}_f$ be the ideal of finite subsets of a space (X, τ). Since $\{x\} \in \mathscr{I}_f$ for each $x \in X$, $A^\star = A^{d^\star}$. As we have already observed that $A^\star = A^\omega$, we conclude that ω-accumulation points of A in (X, τ) are precisely the limit points

of A in $(X, \tau^\star(\mathscr{I}_f))$. Since $(X, \tau^\star)$ is T_1 $(\{x\} \in \mathscr{I}_f \to \{x\}^\star = \phi \to \mathrm{cl}^\star\{x\} = \{x\})$ and the concepts of ω-accumulation and limit point coincide in T_1 spaces, we have that the sets of ω-accumulation points of $A \subseteq X$ in (X, τ) and $(X, \tau^\star)$ coincide. Note that, $A^\star(\mathscr{I}_f) = A^d$ iff $\tau = \tau^\star(\mathscr{I}_f)$ iff (X, τ) is T_1.

In Example 2.4 we exhibited a nearly discrete space and noticed that $\tau^\star(\mathscr{I}_f)$ is discrete. If we start with the indiscrete space (X, ψ), then $A^\star(\mathscr{I}_f) = X$ if $A \notin \mathscr{I}_f$, and $A^\star(\mathscr{I}_f) = \phi$ if $A \in \mathscr{I}_f$. Therefore, $\mathrm{cl}^\star(A) = X$ if $A \notin \mathscr{I}_f$ and $\mathrm{cl}^\star(A) = A$ if $A \in \mathscr{I}_f$. This means that $\psi^\star(\mathscr{I}_f)$ is the well-known *cofinite* topology.

Example 2.6. Let $\mathscr{I}_c$ be the ideal of countable subsets of a space (X, τ). We have $A^\star = A^{d^\star}$ and since $A^\star$, as mentioned earlier, is the set of condensation points of A, we conclude that condensation points of A in (X, τ) are precisely limit points of A in $(X, \tau^\star(\mathscr{I}_c))$.

If we start again with the indiscrete space (X, ψ), it is easy to see that $\psi^\star(\mathscr{I}_c)$ is the *cocountable* topology. Let us consider another familiar example. Let (R, τ) be the real line with the usual topology. Then the topology $\tau^\star(\mathscr{I}_c)$ is nothing other than the *countable complement extension topology* [**29**, Example 63]. Note that $\tau^\star(\mathscr{I}_c)$ is the smallest topology generated by $\tau \cup \psi^\star(\mathscr{I}_c)$; i.e., the supremum $\tau \vee \psi^\star(\mathscr{I}_c)$ of τ and $\psi^\star(\mathscr{I}_c)$.

Now, it seems that topologies generated by ideals are useful in constructing counterexamples. We state an important basic property of such topologies that justifies the previous remark. Recall that a set A in a space (X, τ) is *closed* and *discrete* iff $A^d = \phi$.

LEMMA 2.7. *Let (X, τ) be a space with an ideal $\mathscr{I}$ on X. If $I \in \mathscr{I}$, then I is closed and discrete in $(X, \tau^\star)$.*

Proof. $I \in \mathscr{I} \to I^\star = \phi \to I^{d^\star} = \phi$. ■

Example 2.8. Let (X, τ) be a space. Then it is easy to check that $\mathscr{I}_{cd} = \{A \subseteq X : A^d = \phi\}$ is an ideal on X and that $A^d \subseteq A^\star$. Therefore $\tau^\star = \tau$. (We remark that in general $\tau^\star(\mathscr{I}) = \tau$ iff every $I \in \mathscr{I}$ is closed in (X, τ).) Moreover, Lemma 2.7 implies that I_{cd} is the largest ideal on X with the property $\tau^\star = \tau$. Finally, $A^\star = A^d$ iff (X, τ) is T_1.

Example 2.9. Let (X, τ) be a space and $A \subseteq X$. Observe that $\mathscr{I}(A) = \{B \subseteq X : B \subseteq A\}$ is an ideal on X. We offer two examples of topologies generated by $\mathscr{I}(A)$.

Let (X, ψ) be the indiscrete space and $p \in X$. Then $\mathscr{I}(X - \{p\}) = \{A \subseteq X : p \notin A\}$ generates a simple topology $\psi^\star = \{A \subseteq X; p \in A\} \cup \{\phi\}$ known as the *particular point topology* [**29**, Examples 8–12].

For a more "sophisticated" example of a topology generated by an ideal, let (R, τ) be the real line with the usual topology and $A = \{1/n : n \in N\}$. Then $\tau^\star(\mathscr{I}(A))$ is known as *Smirnov's topology* [**29**, Example 64].

In our next example we consider the important ideal of nowhere dense sets.

Example 2.10. Let (X, τ) be a space and let $\mathscr{I}_n = \{A \subseteq X : \mathrm{Int}(\mathrm{cl}(A)) = \phi\}$. $\mathscr{I}_n$ is the collection of *nowhere dense* subsets of (X, τ). It is a simple task to show that $\mathscr{I}_n$ is an ideal on X and that this important ideal has a nice property [**30**]; namely the local function of a set $A \subseteq X$ can be given "explicitly":

$$A^\star(\mathscr{I}_n) = \mathrm{cl}(\mathrm{Int}(\mathrm{cl}(A))).$$

Here we have that

$$\mathrm{cl}^\star(A) = A \cup \mathrm{cl}(\mathrm{Int}(\mathrm{cl}(A)))$$

which gives a nontrivial finer topology than τ in general. In [**21**] a topology τ^α has been introduced by defining its open sets to be sets $A \subseteq X$ with $A \subseteq \mathrm{Int}(\mathrm{cl}(\mathrm{Int}(A)))$. Since closed sets in (X, τ^α) are sets $A \subseteq X$ with $\mathrm{cl}(\mathrm{Int}(\mathrm{cl}(A))) \subseteq A$; i.e., $A^\star(\mathscr{I}_n, \tau) \subseteq A$, we have that $\tau^\alpha = \tau^\star(\mathscr{I}_n, \tau)$.

We also point out that in [**7**] it was shown that a class of topologies $[\tau]$ on X (τ is a topology on X) defined as $\{\sigma : \sigma$ is a topology on X and for every $A \subseteq X$, $A \subseteq \mathrm{cl}_\tau(\mathrm{Int}_\tau(A))$ iff $A \subseteq \mathrm{cl}_\sigma(\mathrm{Int}_\sigma(A))\}$ possesses a finest topology for the class, denoted by $F(\tau)$. It is left to the reader to establish that $F(\tau) = \tau^\star(\mathscr{I}_n, \tau)$.

Our next example is likely the most important one and it was a primary motivation for the introduction of a local function. First we need the following definition.

Definition 2.11. An ideal $\mathscr{I}$ is said to be a *σ-ideal* if it is *countably additive*; that is, if $I_n \in \mathscr{I}$ for each $n \in N$, then $\bigcup\{I_n : n \in N\} \in \mathscr{I}$.

Example 2.12. A subset of a space is said to be *meager* (or of *first category*) if it is a countable union of nowhere dense sets. Clearly, the collection of all meager subsets of a space is a σ-ideal. We denote this ideal by $\mathscr{I}_m$. Since non-meager sets are also called the sets of second category, the points of $A^\star(\mathscr{I}_m)$, where A is a set in a space, are the points of second category of A [**17**]. In [**3**] Blumberg defines a point x in a space to be "inexhaustibly approached" by a set A if $x \in A^\star(\mathscr{I}_m)$.

In the case of $\mathscr{I}_n$, the ideal of nowhere dense sets in a space (X, τ), we established that $A^\star(\mathscr{I}_n) = \mathrm{cl}(\mathrm{Int}(\mathrm{cl}(A)))$. We point out that $A^\star(\mathscr{I}_m)$ has the following significant property: $A^\star = \mathrm{cl}(\mathrm{Int}(A^\star))$. (The reader is invited to show this by using $A^\star(\mathscr{I}_n) = \mathrm{cl}(\mathrm{Int}(\mathrm{cl}(A)))$, Theorem 4.12 and Theorem 4.6, (b).)

3. The Open Sets of $\tau^\star$. We have defined $\tau^\star$ in terms of the closure operator $\mathrm{cl}^\star(A) = A \cup A^\star$. However, a simple basis for the open sets of $\tau^\star$ can be described.

Let (X, τ) be a space, $\mathscr{I}$ an ideal on X, and observe that A is $\tau^\star$-closed iff $A^\star \subseteq A$. Now we have $U \in \tau^\star$ iff $X - U$ is $\tau^\star$-closed iff $(X - U)^\star \subseteq X - U$ iff $U \subseteq X - (X - U)^\star$. Therefore $x \in U \to x \notin (X - U)^\star \to$ there exists a $V \in N(x)$ such that $V \cap (X - U) \in \mathscr{I}$. Now let $I = V \cap (X - U)$ and we have $x \in V - I \subseteq U$, where $V \in \tau$ and $I \in \mathscr{I}$. We shall denote

$$\beta(\mathscr{I}, \tau) = \{V - I : V \in \tau, I \in \mathscr{I}\},$$

and, following our convention, we will simply write $\beta(\mathscr{I})$ or β when no ambiguity is present. These remarks suggest the following result.

THEOREM 3.1 [**30**]. *Let (X, τ) be a space and $\mathscr{I}$ an ideal on X. Then β is a basis for $\tau^\star$.*

Proof. In view of the preceding remarks, we need only show that β is closed under finite intersections. This follows immediately from De Morgan's laws and the finite additivity of $\mathscr{I}$. ■

If we now return to Example 2.6 we see that the remark $\tau^\star(\mathscr{I}_c) = \tau \vee \psi^\star(\mathscr{I}_c)$ is not at all surprising in the light of Theorem 3.1. We have the following theorem.

THEOREM 3.2. *Let* (X, τ) *be a space and* $\mathscr{I}$ *an ideal on* X. *Then* $\tau^\star(\mathscr{I}) = \tau \vee \psi^\star(\mathscr{I})$ *where* ψ *denotes the indiscrete topology.* ■

A question which should occur to us at this point is: Given a space (X, τ) and an ideal $\mathscr{I}$ on X, we can form $[\tau^\star(\mathscr{I})]^\star(\mathscr{I}) = \tau^{\star\star}$. We know that $\tau^{\star\star}$ is finer than $\tau^\star$, but is it ever true that $\tau^{\star\star}$ is strictly finer than $\tau^\star$? The following theorem will lead to an answer of this question in the negative. But first, observe that if $\mathscr{I}$ and $\mathscr{J}$ are ideals on X, then $\mathscr{I} \cap \mathscr{J}$ and $\mathscr{I} \vee \mathscr{J} = \{I \cup J : I \in \mathscr{I} \text{ and } J \in \mathscr{J}\}$ are also ideals.

THEOREM 3.3. *Let* (X, τ) *be a space with* $\mathscr{I}$ *and* $\mathscr{J}$ *ideals on* X, *and* $A \subseteq X$. *Then*
(a) **[31]** $A^\star(\mathscr{I} \cap \mathscr{J}) = A^\star(\mathscr{I}) \cup A^\star(\mathscr{J})$.
(b) $A^\star(\mathscr{I} \vee \mathscr{J}, \tau) = A^\star(\mathscr{I}, \tau^\star(\mathscr{J})) \cap A^\star(\mathscr{J}, \tau^\star(\mathscr{I}))$.

Proof. (a) is clear. To show (b) assume that $x \notin A^\star(\mathscr{I} \vee \mathscr{J}, \tau)$. Then there exists a $U \in \mathscr{N}(x)$ such that $U \cap A \in \mathscr{I} \vee \mathscr{J}$. Let $I \in \mathscr{I}$ and $J \in \mathscr{J}$ such that $U \cap A = I \cup J$. Because of the heredity of $\mathscr{I}$, we may assume $I \cap J = \phi$. Thus we have $(U \cap A) - I = J$ and $(U \cap A) - J = I \to (U - I) \cap A = J \in \mathscr{J}$ and $(U - J) \cap A = I \in \mathscr{I} \to x \notin A^\star(\mathscr{J}, \tau^\star(\mathscr{I}))$ or $x \notin A^\star(\mathscr{I}, \tau^\star(\mathscr{J}))$ (x could be in I or J, but not both). We have shown
(1) $A^\star(\mathscr{J}, \tau^\star(\mathscr{I})) \cap A^\star(\mathscr{I}, \tau^\star(\mathscr{J})) \subseteq A^\star(\mathscr{I} \vee \mathscr{J}, \tau)$.
Now assume $x \notin A^\star(\mathscr{I}, \tau^\star(\mathscr{J}))$. This implies there exist $U \in \mathscr{N}(x)$ and $J \in \mathscr{J}$ such that $(U - J) \cap A \in \mathscr{I}$. We may assume, because of the heredity of $\mathscr{J}$, that $J \subseteq A$. Now define $I = (U - J) \cap A$, and we have $U \cap A = I \cup J \in \mathscr{I} \vee \mathscr{J} \to x \notin A^\star(\mathscr{I} \vee \mathscr{J}, \tau)$. We have shown
(2) $A^\star(\mathscr{I} \vee \mathscr{J}, \tau) \subseteq A^\star(\mathscr{I}, \tau^\star(\mathscr{J}))$.
Similarly, we have that
(3) $A^\star(\mathscr{I} \vee \mathscr{J}, \tau) \subseteq A^\star(\mathscr{J}, \tau^\star(\mathscr{I}))$.
(1), (2), and (3) establish the result. ■

By taking $\mathscr{I} = \mathscr{J}$ in the above theorem, the following corollary answers the question about the relationship between $\tau^\star$ and $\tau^{\star\star}$.

COROLLARY 3.4. *Let* (X, τ) *be a space and* $\mathscr{I}$ *an ideal on* X. *Then* $A^\star(\mathscr{I}, \tau) = A^\star(\mathscr{I}, \tau^\star)$ *and hence* $\tau^\star = \tau^{\star\star}$ (*cf.* **[22]**). ■

Note that it is possible to establish Corollary 3.4 by using Lemma 2.7 and remarks in Example 2.8.

Theorem 3.3 also gives a straightforward proof to the following result.

COROLLARY 3.5. *Let* (X, τ) *be a space with* $\mathscr{I}$ *and* $\mathscr{J}$ *ideals on* X. *Then*
(a) **[26]** $\tau^\star(\mathscr{I} \vee \mathscr{J}) = [\tau^\star(\mathscr{J})]^\star(\mathscr{I}) = [\tau^\star(\mathscr{I})]^\star(\mathscr{J})$
(b) $\tau^\star(\mathscr{I} \vee \mathscr{J}) = \tau^\star(\mathscr{I}) \vee \tau^\star(\mathscr{J})$
(c) **[26]** $\tau^\star(\mathscr{I} \cap \mathscr{J}) = \tau^\star(\mathscr{I}) \cap \tau^\star(\mathscr{J})$.

Proof. (a) follows from Theorem 3.3, (b). To show (b) we have $\tau^\star(\mathscr{I} \vee \mathscr{J}) = [\tau^\star(\mathscr{I})]^\star(\mathscr{J})$ by (a) and hence $\tau^\star(\mathscr{I} \vee \mathscr{J}) = \tau^\star(\mathscr{I}) \vee \psi^\star(\mathscr{J})$ by Theorem 3.2. Since $\tau^\star(\mathscr{I}) \vee \psi^\star(\mathscr{J}) = \tau^\star(\mathscr{I}) \vee \tau \vee \psi^\star(\mathscr{J}) = \tau^\star(\mathscr{I}) \vee \tau^\star(\mathscr{J})$, by Theorem 3.2, we have that (b) follows. (c) follows from Theorem 3.3, (a). ■

We have shown that for a given space (X, τ) and ideal $\mathscr{I}$ on X, β is a basis for $\tau^\star$. If β is itself a topology, then we have $\beta = \tau^\star$ (as in the case of the cocompact

and cocountable topologies) and all the open sets of $\tau^\star$ are of the *simple form* $V - I$ where $V \in \tau$ and $I \in \mathscr{I}$. The following example shows that β is not a topology in general.

Example 3.6. Let (X, τ) be the real numbers with the "left-ray" topology, i.e., $\tau = \{(-\infty, a): a \in X\} \cup \{X, \varnothing\}$. Let $\mathscr{I}_f$ be the ideal of finite subsets of X. Consider the collection of subsets of β defined as $U_n - I_n$ where $U_n = (-\infty, n + 1/2)$ and $I_n = \{1, 2, \ldots, n\}$. Then $\cup\{U_n - I_n: n \in N\} = X - N$ which is not in β. Thus β is not closed under arbitrary unions, and hence is not a topology.

In the following section, we consider a natural condition relating τ and $\mathscr{I}$ that will guarantee β is a topology and hence all sets in $\tau^\star$ will be of simple form.

4. Compatibility of τ with $\mathscr{I}$. Let us begin with the following definition.

Definition 4.1 [**22**]. Let (X, τ) be a space with $\mathscr{I}$ an ideal on X. We say the topology τ is *compatible* with the ideal $\mathscr{I}$, denoted $\tau \sim \mathscr{I}$, if the following holds for every $A \subseteq X$: if for every $x \in A$ there exists a $U \in \mathscr{N}(x)$ such that $U \cap A \in \mathscr{I}$, then $A \in \mathscr{I}$.

Compatible ideals are called "supercompact" in [**31**] and "adherence ideals" in [**30**].

Since ideals are usually defined as a collection of subsets of X having some property $\mathscr{P}$, the condition $\tau \sim \mathscr{I}$ then means that whenever a subset has the property $\mathscr{P}$ "locally," it necessarily has the property "globally." This condition may appear to be quite restrictive upon first consideration. However, the condition holds in many useful, important, and quite general situations. The next theorem gives sufficient conditions for $\tau \sim \mathscr{I}$, but first we need some preliminary definitions.

A space is said to be *hereditarily Lindelöf* if every subspace has the Lindelöf property. Second countable spaces are hereditarily Lindelöf but the converse is not true (consider $(R, \psi^\star(\mathscr{I}_c))$ or $(R, \tau^\star(\mathscr{I}_c))$ in Example 2.6).

THEOREM 4.2. *Let (X, τ) be a hereditarily Lindelöf space and $\mathscr{I}$ a σ-ideal on X, then $\tau \sim \mathscr{I}$.*

Proof. Let $A \subseteq X$ and assume that for every $x \in A$, there exists a $U_x \in \mathscr{N}(x)$ such that $U_x \cap A \in \mathscr{I}$. Now $\{U_x \cap A: x \in A\}$ is an open cover of A (relative to the subspace A) and hence has a countable subcover $\{V_n \cap A: n \in N\}$. Since $\mathscr{I}$ is a σ-ideal, $A = \cup\{V_n \cap A: n \in N\}$ is in $\mathscr{I}$. ■

Example 4.3. Let (R, τ) be the real numbers with the usual topology and $\mathscr{I}$ the σ-ideal of subsets of Lebesgue measure zero. Clearly, $\tau \sim \mathscr{I}$. The topology $\tau^\star(\mathscr{I})$ has the following nice property: *$\tau^\star$-Borel sets are precisely the Lebesgue measurable sets of τ*. For further details see [**27**].

The significance of compatibility is shown by the following theorem.

THEOREM 4.4 [**22**]. *Let (X, τ) be a space with $\mathscr{I}$ an ideal on X. If $\tau \sim \mathscr{I}$, then β is a topology (and, hence, $\beta = \tau^\star$).*

We will return to the proof of this theorem after establishing some preliminary results.

THEOREM 4.5. *Let (X, τ) be a space with $\mathscr{I}$ an ideal on X. Then the following are equivalent*:

(a) $\tau \sim \mathscr{I}$.

(b) [**31**] *If A has a cover of open sets each of whose intersection with A is in $\mathscr{I}$, then A is in $\mathscr{I}$.*

(c) [**31**] *For every $A \subseteq X$, $A \cap A^\star = \phi \to A \in \mathscr{I}$.*

(d) [**31**] *For every $A \subseteq X$, $A - A^\star \in \mathscr{I}$.*

(e) *For every $\tau^\star$-closed subset A, $A - A^\star \in \mathscr{I}$.*

(f) *For every $A \subseteq X$, if A contains no nonempty subset B with $B \subseteq B^\star$, then $A \in \mathscr{I}$.*

Proof. (c) → (d); Let $A \subseteq X$, then observe that $(A - A^\star \cap (A - A^\star)^\star = \phi$, and apply (c).

(e) → (a). Let $A \subseteq X$ and assume that for every $x \in A$, there exists a $U \in \mathscr{N}(x)$ such that $U \cap A \in \mathscr{I}$. Then $A \cap A^\star = \phi$ and since $A \cup A^\star$ is $\tau^\star$-closed, we have $(A \cup A^\star) - (A \cup A^\star)^\star \in \mathscr{I}$, and $(A \cup A^\star) - (A \cup A^\star)^\star = (A \cup A^\star) - (A^\star \cup A^{\star\star}) = (A \cup A^\star) - A^\star = A$. Therefore $A \in \mathscr{I}$.

(d) → (f) Let $A \subseteq X$ and assume that A has no nonempty subset B with $B \subseteq B^\star$. Since $A - A^\star \in \mathscr{I}$, $A \cap A^\star \subseteq (A \cap A^\star)^\star$ (See Theorem 4.6 (c)) and hence $A \cap A^\star = \phi$. Therefore $A = A - A^\star$ and $A \in \mathscr{I}$.

(f) → (d). Let $A \subseteq X$. Since $(A - A^\star) \cap (A - A^\star)^\star = \phi$, $A - A^\star$ contains no nonempty subset B with $B \subseteq B^\star$. Hence $A - A^\star \in \mathscr{I}$. ■

THEOREM 4.6. *Let (X, τ) be a space with $\mathscr{I}$ an ideal on X. Then the following are equivalent, and are implied by $\tau \sim \mathscr{I}$.*

(a) *For every $A \subseteq X$, $A \cap A^\star = \phi \to A^\star = \phi$.*

(b) *For every $A \subseteq X$, $(A - A^\star)^\star = \phi$.*

(c) *For every $A \subseteq X$, $(A \cap A^\star)^\star = A^\star$.* ■

Theorem 4.4 will be established as a corollary to the following theorem.

THEOREM 4.7 [**22**]. *Let (X, τ) be a space and $\mathscr{I}$ an ideal on X compatible with τ. A set is closed with respect to $\tau^\star$ iff it is the union of a set which is closed with respect to τ and a set in $\mathscr{I}$.*

Proof. Let A be $\tau^\star$-closed, then $A^\star \subseteq A \to A = (A - A^\star) \cup A^\star$. Now $A - A^\star \in \mathscr{I}$ by Theorem 4.5, (d) and $A^\star$ is τ-closed by Theorem 2.3, (c).

Conversely, if $A = B \cup I$ where B is τ-closed and $I \in \mathscr{I}$, then $A^\star = B^\star \cup I^\star = B^\star \subseteq \mathrm{cl}(B) = B \subseteq A$. Thus $A^\star \subseteq A$ and A is $\tau^\star$-closed. ■

Theorem 4.4 is now immediate since every closed set in $\tau^\star$ (under the hypothesis of Theorem 4.4 that $\tau \sim \mathscr{I}$) is of the simple form $B \cup I$ with B τ-closed and $I \in \mathscr{I}$. Thus every $\tau^\star$-open set will be the complement of such a set, and hence have the simple form $U - I$ where $U \in \tau$ and $I \in \mathscr{I}$.

While compatibility is a sufficient condition for β to be a topology and $\tau^\star = \beta$, the following example shows that compatibility is not a necessary condition. The example also shows that the equivalent conditions of Theorem 4.6 are weaker than compatibility.

Example 4.8. Let (X, τ) be the real numbers with the discrete topology and $\mathscr{I}_c$ the σ-ideal of countable subsets of X. Then we have $\tau = \beta = \tau^\star$ but τ is not

compatible with $\mathscr{I}_c$. For example, X is certainly locally countable at each point but X is not countable.

In this space, $A^\star = \phi$ for every $A \subseteq X$. Thus condition (a) of Theorem 4.6 holds.

Remark 4.9. The previous example points out an error in [**31**] where the author claims condition (a) of Theorem 4.6 is equivalent to compatibility. We point out, however, that if an ideal satisfies "$A^\star = \phi \to A \in \mathscr{I}$" for every $A \subseteq X$ (called "compactness" in [**31**]) then the conditions of Theorem 4.6 are equivalent to compatibility. Instead of "$\mathscr{I}$ is compact" we will say that τ is *weakly compatible with* $\mathscr{I}$, denoted $\tau \overset{\omega}{\sim} \mathscr{I}$. To see how strong $\overset{\omega}{\sim}$ is compared with $\sim$, note that in the case of $\mathscr{I}_f$, $\tau \overset{\omega}{\sim} \mathscr{I}_f$ iff (X, τ) is countably compact [**31**]. It is claimed in [**30**] that for T_1 spaces, τ is never compatible with $\mathscr{I}_f$, except in the trivial case when X is finite and hence $\mathscr{I}_f = \mathscr{P}(X)$. This is not correct, however, since $\tau \sim \mathscr{I}_f$ iff (X, τ) is hereditarily compact. The situation is more interesting in the case of $\mathscr{I}_c$. We obtain that $\tau \overset{\omega}{\sim} \mathscr{I}_c$ iff $(X, \tau^\star)$ is κ_1-compact (a space is *κ_1-compact* iff every uncountable set in a space has at least one limit point). Note that (X, τ) being Lindelöf $\to \tau \overset{\omega}{\sim} \mathscr{I}_c \to (X, \tau)$ is κ_1-compact.

The case $\tau \sim \mathscr{I}_c$ is important since it gives the following converse of Theorem 4.2.

THEOREM 4.10. *A space (X, τ) is hereditarily Lindelöf iff $\tau \sim \mathscr{I}_c$.*

Proof. We have only to show that the condition is sufficient. Suppose that (X, τ) is not hereditarily Lindelöf. Then there exists an uncountable set $A \subseteq X$ such that for each $x \in A$ there exists a $U \in \mathscr{N}(x)$ with $U \cap A$ countable [**18**]. This simply means there is an $A \notin \mathscr{I}_c$ with $A \cap A^\star(\mathscr{I}_c) = \phi$. This contradicts $\tau \sim \mathscr{I}_c$ (Theorem 4.5, (c)). ■

Let us return now to Example 2.10 to establish that all the closed (and open) sets in $(X, \tau^\star(\mathscr{I}_n))$ are of simple form.

THEOREM 4.11. *Let (X, τ) be a space and $\mathscr{I}_n$ the ideal of nowhere dense subsets of X. Then $\tau \sim \mathscr{I}_n$.*

Proof. We will show that condition (c) of Theorem 4.5 is satisfied. Assume that $A \cap A^\star(\mathscr{I}_n) = \phi \to A \cap \mathrm{cl}(\mathrm{Int}(\mathrm{cl}(A))) = \phi \to A \cap \mathrm{Int}(\mathrm{cl}(\mathrm{Int}(\mathrm{cl}(A)))) = \phi \to A \cap \mathrm{Int}(\mathrm{cl}(A)) = \phi \to \mathrm{cl}(A) \cap \mathrm{Int}(\mathrm{cl}(A)) = \phi \to \mathrm{Int}(\mathrm{cl}(A)) = \phi$. ■

A more difficult task is to establish that $\tau \sim \mathscr{I}_m$. For the proof, we refer the reader to [**23**] (Theorem 16.1 of [**23**] is equivalent to $\tau \sim \mathscr{I}_m$).

THEOREM 4.12. (*Banach Category Theorem*). *Let (X, τ) be a space and $\mathscr{I}_m$ the ideal of meager subsets, then $\tau \sim \mathscr{I}_m$.* ■

Remark 4.13. We conclude this section by noticing that compatibility and weak compatibility are properties shared by τ and $\tau^\star$, namely, $\tau \sim \mathscr{I}$ (resp., $\tau \overset{\omega}{\sim} \mathscr{I}$) iff $\tau^\star \sim \mathscr{I}$ (resp., $\tau^\star \overset{\omega}{\sim} \mathscr{I}$) (cf. [**22**]).

5. Ideals containing $\mathscr{I}_f$. We have observed in §2 that if $\{x\} \in \mathscr{I}$ for each $x \in X$, then $A^\star = A^{d^\star}$ for every $A \subseteq X$. Since the converse is obvious, we state the following theorem, for reference.

THEOREM 5.1. *Let (X, τ) be a space with $\mathscr{I}$ an ideal on X. Then $\{x\} \in \mathscr{I}$ for each $x \in X$ iff $A^\star = A^{d^\star}$ for every $A \subseteq X$.* ■

Now, recall that a subset A of a space (X, τ) is *dense-in-itself* if $A \subseteq A^d$; that is, A is without *isolated points*. A set which is dense-in-itself and closed is said to be *perfect*. We need this concept to state an old result proved by Cantor and Bendixson independently in 1883 for subsets of the real line.

THEOREM 5.2 (*Cantor-Bendixson*). *A second countable* (*moreover, hereditarily Lindelöf*) *space can be represented as the union of two sets, one of which is perfect and the other countable.* ■

Let us see now the proof of this theorem in a general setting.

THEOREM 5.3 [**9**]. *Let (X, τ) be a space with an ideal $\mathcal{I}$ on X such that $\tau \sim \mathcal{I}$ and $\{x\} \in \mathcal{I}$ for each $x \in X$. If a set $A \subseteq X$ is closed with respect to $\tau^\star$, then A is the union of a set which is perfect with respect to τ and a set in $\mathcal{I}$.*

Proof. Let $A \subseteq X$ be closed with respect to $\tau^\star$, then by Theorem 4.5 (d), $A = A^\star \cup I$ with $I \in \mathcal{I}$ (and $A^\star \cap I = \phi$). Since $A^\star - A^{\star\star} \subseteq (A - A^\star)^\star$ by Theorem 2.3, (f) and $(A - A^\star)^\star = \phi$ by Theorem 4.6 (b), we have $A^\star = A^{\star\star}$. By Theorem 5.1, $A^\star = (A^\star)^{d^\star}$ and consequently $A^\star \subseteq (A^\star)^d$. So, $A^\star$ is perfect with respect to τ and the result follows. ■

To see that Theorem 5.3 generalizes the Cantor-Bendixson Theorem, set $A = X$ and $\mathcal{I} = \mathcal{I}_c$ in Theorem 5.3.

Recall that a set A is *discrete* if $A \cap A^d = \phi$. The collection of all discrete subsets of a space (X, τ) fails to form an ideal on X; however, there is a larger collection that does form an important ideal. A set $A \subseteq X$ is said to be *scattered* if A contains no nonempty dense-in-itself subset. The proof that the collection of scattered sets in a T_1 space is finitely additive is in [**17**]. For further properties of this ideal, denoted by $\mathcal{I}_s$, see [**30**] and [**24**]. We point out that discrete sets are scattered while the converse is not true in general. In Lemma 2.7 we established that sets in $\mathcal{I}$ are closed and discrete in $(X, \tau^\star(\mathcal{I}))$ and hence are scattered in $(X, \tau^\star(\mathcal{I}))$. What is a condition that forces scattered sets of $(X, \tau^\star)$ to be in $\mathcal{I}$? Obviously, $\{x\}$ is discrete, and hence scattered in $(X, \tau^\star)$ for each $x \in X$, and is not necessarily in $\mathcal{I}$. Consequently, we need $\{x\} \in \mathcal{I}$ for each $x \in X$. The additional condition is, of course, compatibility.

THEOREM 5.4. *Let (X, τ) be a space and $\mathcal{I}$ an ideal on X. The following are equivalent.*

(a) *$\tau \sim \mathcal{I}$ and $\{x\} \in \mathcal{I}$ for each $x \in X$.*

(b) *Scattered sets in $(X, \tau^\star)$ are in $\mathcal{I}$.*

(c) *Discrete sets in $(X, \tau^\star)$ are in $\mathcal{I}$.*

Proof. (a) → (b). Let $A \subseteq X$, then $A^\star = A^{d^\star}$ by Theorem 5.1 and (b) follows by Theorem 4.5, (f).

(b) → (c). Clear.

(c) → (a). Obviously, $\{x\} \in \mathcal{I}$ for each $x \in X$. Let $A \subseteq X$, then $A - A^\star$ is discrete in $(X, \tau^\star)$ by Theorem 5.1 since $(A - A^\star) \cap (A - A^\star)^\star = \phi$. Now, Theorem 4.5, (c) implies $\tau \sim \mathcal{I}$. ■

Combining Theorem 5.4, Remark 4.9, and Remark 4.13, we have that a space (X,τ) is hereditarily compact iff $(X, \tau^\star(\mathcal{I}_f))$ is hereditarily compact iff $(X, \tau^\star(\mathcal{I}_f))$ has no discrete infinite subspace.

In a similar way, we obtain the following result.

COROLLARY 5.5. *Let* (X, τ) *be a space and* $\mathscr{I}_c$ *be the ideal of countable subsets of* X. *The following are equivalent.*

(a) (X, τ) *is hereditarily Lindelöf.*

(b) $(X, \tau^\star(\mathscr{I}_c))$ *is hereditarily Lindelöf.*

(c) *If* I *is scattered in* $(X, \tau^\star)$, *then* $I \in \mathscr{I}_c$.

(d) *If* I *is discrete in* $(X, \tau^\star)$, *then* $I \in \mathscr{I}_c$. ■

As a consequence of Corollary 5.5, we have for instance, that $(R, \tau^\star(\mathscr{I}_c))$ (Example 2.6) is a hereditarily Lindelöf space and that scattered sets in $(R, \tau^\star(\mathscr{I}_c))$ are precisely the countable subsets of R. As an application of Corollary 5.5 we show that T_1 P-spaces with discrete subspaces countable are hereditarily Lindelöf (cf. [**25**]). First, recall that a space in which the intersection of a countable collection of open sets is open, is called a P-space. It is clear that in a T_1 P-space (X, τ), $\mathscr{I}_c \subseteq \mathscr{I}_{cd}$. Therefore $\tau^\star(\mathscr{I}_c) = \tau^\star(\mathscr{I}_{cd}) = \tau$. If we assume that discrete subspaces of (X, τ) are countable, (X, τ) is hereditarily Lindelöf by Corollary 5.5, (d).

6. Spaces for which $X = X^\star$. An open subset U of a space (X, τ) is said to be *regular open* if $U = \text{Int}(\text{cl}(U))$. The regular open sets in (X, τ) form a basis for a new topology on X, known as the *semiregularization* of τ, denoted τ_s. The topology τ_s is coarser than τ, and τ is said to be *semiregular* if $\tau = \tau_s$.

Given a space (X, τ) and an ideal $\mathscr{I}$ on X, a hypothesis used by Hayashi [**12**] is that $X = X^\star$ and a hypothesis used by Samuels [**26**] is $\tau \cap \mathscr{I} = \{\phi\}$. These equivalent conditions imply $\tau_s = (\tau^\star)_s$, as we shall see.

THEOREM 6.1. *Let* (X, τ) *be a space with* $\mathscr{I}$ *an ideal on* X. *The following are equivalent.*

(a) $X = X^\star$,

(b) $\tau \cap \mathscr{I} = \{\phi\}$,

(c) *If* $I \in \mathscr{I}$, *then* $\text{Int}(I) = \phi$, *and*

(d) *For every* $U \in \tau$, $U \subseteq U^\star$. ■

Since $X^\star(\mathscr{I}, \tau) = X^\star(\mathscr{I}, \tau^\star)$ by Corollary 3.4, we may replace τ by $\tau^\star$ in (b) of Theorem 6.1, "$\text{Int}(A) = \phi$" by "$\text{Int}^\star(A) = \phi$" in (c) and "$U \in \tau$" by "$U \in \tau^\star$" in (d).

The condition $X = X^\star$ is satisfied in the case of the ideal $\mathscr{I}_n$ of nowhere dense sets in (X, τ). Also, it is clear that $X = X^\star$ in the case of $(R, \tau^\star(\mathscr{I}_c))$ (Example 2.6). To see the importance of $X = X^\star$ consider the ideal of meager sets in (X, τ). Recall that a space (X, τ) is said to a *Baire* space if the intersection of every countable family of open dense sets in (X, τ) is dense. Now, it is not difficult to see that (X, τ) is a Baire space iff $X = X^\star(\mathscr{I}_m)$.

In order to prove the promised result that (X, τ) and $(X, \tau^\star)$ have the same semiregularizations if $X = X^\star$, we need the following folklore lemma whose proof we omit.

LEMMA 6.3. *Let* τ *and* σ *be topologies on* X *and* $\tau \subseteq \sigma$. *If* $\text{cl}_\tau(V) = \text{cl}_\sigma(V)$ *for every* $V \in \sigma$, *then* $\tau_s = \sigma_s$. ■

THEOREM 6.4. *Let* (X, τ) *be a space with an ideal* $\mathscr{I}$ *on* X *such that* $X = X^\star$. *Then* $\tau_s = (\tau^\star)_s$.

Proof. Let $V \in \tau^\star$. From remarks following Theorem 6.1 it is clear that, $V \subseteq V^\star$ and hence $\mathrm{cl}(V) \subseteq V^\star$. This gives $\mathrm{cl}(V) \subseteq \mathrm{cl}^\star(V)$ and the result follows. ■

COROLLARY 6.5. *Let (X, τ) be a space with an ideal $\mathscr{I}$ on X such that $X = X^\star$. If $(X, \tau^\star)$ is semiregular, then $\tau = \tau^\star$.*

Proof. Observe that $\tau^\star = (\tau^\star)_s = \tau_s \subseteq \tau \subseteq \tau^\star$. ■

Let (X, τ) be a space with an ideal $\mathscr{I}$ on X and $A \subseteq X$. Then $\mathscr{I}_A = \{A \cap I : I \in \mathscr{I}\}$ is obviously an ideal on A. The following easily proved lemma will be useful in generalizing Theorem 6.4.

LEMMA 6.6. *Let (X, τ) be a space with an ideal $\mathscr{I}$ on X and $A \subseteq X$. Then $(\tau/A)^\star(\mathscr{I}_A) = \tau^\star(\mathscr{I})/A$.* ■

THEOREM 6.7. *Let (X, τ) be a space with an ideal $\mathscr{I}$ on X and $A \subseteq X$. If $A \subseteq A^\star$, then $(\tau/A)_s = (\tau^\star/A)_s$.*

Proof. We will show that $(\tau/A) \cap \mathscr{I}_A = \{\phi\}$. Suppose that $U \in \tau/A \cap \mathscr{I}_A$. Then $U \in \mathscr{I}$ and there is a $V \in \tau$ such that $U = V \cap A$. This implies $(V \cap A)^\star = \phi$ and by Theorem 2.3, (g), $V \cap A^\star = \phi$. Since $A \subseteq A^\star$, $U = \phi$. Hence $\tau/A \cap \mathscr{I}_A = \{\phi\}$ and the result follows by Theorem 6.1, Lemma 6.6, and Theorem 6.4. ■

The following easily proved corollaries also give sufficient conditions for the restrictions of τ and $\tau^\star$ on a subset to have the same semiregularizations.

COROLLARY 6.8. *Let (X, τ) be a space with an ideal $\mathscr{I}$ on X and $A \subseteq X$. If $A^\star = A^{\star\star}$, then $(\tau/A^\star)_s = (\tau^\star/A^\star)_s$.* ■

COROLLARY 6.9. *Let (X, τ) be a space with an ideal $\mathscr{I}$ on X such that $X - X^\star \in \mathscr{I}$. Then $(\tau/X^\star)_s = (\tau^\star/X^\star)_s$.* ■

It was observed in [5] that some important topological properties are shared by (X, τ) and (X, τ_s). Some of these properties, so called *semiregular* properties, are: Hausdorffness, property of a space being Urysohn, connectedness, and extremal disconnectedness [5]. The list of semiregular properties is more extensive and it is known that H-closedness, light compactness, pseudocompactness and some other properties are semiregular. The following corollary now follows immediately from Theorem 6.4.

COROLLARY 6.10. *Semiregular properties are shared by (X, τ) and $(X, \tau^\star)$ if $X = X^\star$.* ■

The concept of θ-continuous function [8] is closely related to that of semiregularization. A function $f:(X, \tau) \to (Y, \sigma)$ is said to be *θ-continuous* at a point $x \in X$ if for every $V \in \mathscr{N}_\sigma(f(x))$ there exists a $U \in \mathscr{N}_\tau(x)$ such that $f(\mathrm{cl}(U)) \subseteq \mathrm{cl}(V)$. A function $f: (X, \tau) \to (Y, \sigma)$ is θ-continuous if f is θ-continuous at x for every $x \in X$. The concept of θ-continuity is weaker than that of continuity as the identiy function $i: (X, \tau_s) \to (X, \tau)$ for a nonsemiregular space (X, τ) shows. But if the range of a θ-continuous function is regular, the function is continuous. It is not difficult to show that $f: (X, \tau) \to Y$ is θ-continuous iff $f: (X, \tau_s) \to Y$ is θ-continuous. Hence, the following results.

THEOREM 6.11. *If $X = X^\star$, then $f: (X, \tau) \to Y$ is θ-continuous* iff *$f: (X, \tau^\star) \to Y$ is θ-continuous.* ■

COROLLARY 6.12 [**26**]. *If $X = X^\star$ and Y is regular, then f: $(X, \tau) \to Y$ is continuous iff f: $(X, \tau^\star) \to Y$ is continuous.* ■

7. Some Applications.

A. *Spaces in which Compact Sets are Finite*

The class of spaces in which compact sets are finite has been a subject of some interest in recent years. Clearly, "The general problem is to move as far away from discreteness (as possible) and yet to retain the property that only finite sets be compact" [**28**]. Open problems in this area can be found in [**1**].

Spaces in which compact sets are finite have been called *cf-spaces* in [**19**], *pseudofinite* in [**32**], and recently, *anticompact* in [**1**]. As pointed out in [**1**], an important source of anticompact examples are the MI spaces of Hewitt [**13**]. A space is said to be MI if it is dense-in-itself, Hausdorff, and submaximal in the sense of Bourbaki [**5**], i.e., every dense subset is open. In [**15**] it was shown that MI spaces are anticompact and this result was improved in [**28**] where it was shown that dense-in-themselves Hausdorff spaces with the property that sets with dense interiors are open are anticompact. If we rephrase the last result we have that dense-in-themselves Hausdorff spaces with nowhere dense sets closed are anticompact.

We will show how to obtain an anticompact space by starting with a Hausdorff space and generating a finer topology via an appropriate ideal. Besides $\mathscr{I}_c$ and $\mathscr{I}_s$, the ideal of relatively compact sets $\mathscr{I}_k = \{A \subseteq X: \mathrm{cl}(A) \text{ is compact in } (X, \tau)\}$ will play an important role. Note that a Hausdorff space (X, τ) is anticompact iff $\tau = \tau^\star(\mathscr{I}_k)$. Our next result provides a rich source of anticompact spaces.

THEOREM 7.1. *Let (X, τ) be a Hausdorff space. Then $(X, \tau^\star(\mathscr{I}_c \cap \mathscr{I}_s \cap \mathscr{I}_k))$ is anticompact.*

Proof. Let A be an infinite subset of X. Suppose that A is compact in $(X, \tau^\star)$. Since (X, τ) is Hausdorff, there exists a countable subset B of A which is discrete in (X, τ). It is clear that $B \in \mathscr{I}_c \cap \mathscr{I}_s$. Since A is compact in $(X, \tau^\star)$, A is compact in (X, τ), and hence closed in (X, τ). So, $B \in \mathscr{I}_k$ and consequently, $B \in \mathscr{I}_c \cap \mathscr{I}_s \cap \mathscr{I}_k$. By Lemma 2.7, B is closed and discrete in $(X, \tau^\star)$. This contradicts the assumption. Therefore $(X, \tau^\star)$ possesses no infinite compact set. ■

We point out that $\tau^\star(\mathscr{I}_c \cap \mathscr{I}_s \cap \mathscr{I}_k) = \tau^\star(\mathscr{I}_c) \cap \tau^\star(\mathscr{I}_s) \cap \tau^\star(\mathscr{I}_k)$ (Corollary 3.5, (c)). Therefore these three topologies are anticompact. Clearly, for $\tau^\star(\mathscr{I}_c)$ it is not necessary to assume that (X, τ) is Hausdorff. The fact that $\tau^\star(\mathscr{I}_c)$ is anticompact implies that a space $(X, \tau^\star)$ is anticompact if $\mathscr{I}$ is a σ-ideal on X and $\{x\} \in \mathscr{I}$ for each $x \in X$ [**20**], as well as that T_1 P-spaces are anticompact [**1**].

We remark that first countable T_1 anticompact spaces are discrete and that locally compact Hausdorff anticompact spaces are also discrete. More generally, Hausdorff anticompact k-spaces are discrete [**1**].

We now use Theorem 7.1 to offer a general result.

THEOREM 7.2. *Let (X, τ) be a Hausdorff space with an ideal $\mathscr{I}$ on X. If $\tau \sim \mathscr{I}$ and $\{x\} \in \mathscr{I}$ for each $x \in X$, then $(X, \tau^\star(\mathscr{I}))$ is anticompact.*

Proof. By Theorem 5.4 and Lemma 2.7, $\mathscr{I} = \mathscr{I}_s(\tau^\star)$, where $\mathscr{I}_s(\tau^\star)$ denotes the ideal of scattered sets in $(X, \tau^\star)$. Since $\tau^{\star\star} = \tau^\star$ (Corollary 3.4), the result follows by Theorem 7.1. ■

Now, we return to the MI spaces. Observe that in a dense-in-itself T_1 space scattered sets are nowhere dense [**17**]. As a consequence of Theorem 7.1 we have the following.

COROLLARY 7.3. *Let* (X, τ) *be a dense-in-itself Hausdorff space. Then* $(X, \tau^\star(\mathscr{I}_c \cap \mathscr{I}_n \cap \mathscr{I}_k))$ *is anticompact.* ■

It is left to the reader to establish the previously mentioned result that "dense-in-themselves Hausdorff spaces with nowhere dense sets closed are anticompact" is logically equivalent to the following corollary.

COROLLARY 7.4. *Let* (X, τ) *be a dense-in-itself Hausdorff space. Then* $(X, \tau^\star(\mathscr{I}_n))$ *is anticompact.* ■

B. *Continuity with respect to* $\tau^\star$.

If $f\colon (X, \tau) \to Y$ is a continuous function and $\mathscr{I}$ is an ideal on X, it is clear that $f\colon (X, \tau^\star) \to Y$ is also continuous. The converse is not true as the following example shows [**14**].

Example 7.6. Let (R, τ) be the real line with the usual topology, σ, Smirnov's topology (Example 2.9), and $i\colon (R, \tau) \to (R, \sigma)$ the identity function. Then $i\colon (R, \tau^\star(\mathscr{I}_c)) \to (R, \sigma)$ is continuous since $A = \{1/n : n \in N\} \in \mathscr{I}_c$, while $i\colon (R, \tau) \to (R, \sigma)$ is not. This example also shows that the assumption of regularity of Y in Corollary 7.8 is essential.

THEOREM 7.7. *Let* $f\colon (X, \tau) \to Y$ *be a function and* $\mathscr{I}$ *an ideal on* X. *If* $f\colon (X, \tau^\star) \to Y$ *is* θ*-continuous, then* $f/X^\star\colon (X^\star, \tau/X^\star) \to Y$ *is* θ*-continuous.*

Proof. First note that by Theorem 2.3, (g), $U \cap X^\star = U \cap (U \cap X)^\star = U \cap U^\star$ for every $U \in \tau$. Now let $x \in X^\star$ and $V \in \mathscr{N}(f(x))$. Since $f\colon (X, \tau^\star) \to Y$ is θ-continuous, there exists a basic open set $U - I$ in $(X, \tau^\star)$ containing x, where $U \in \tau$ and $I \in \mathscr{I}$, such that $f(\mathrm{cl}^\star(U - I)) \subseteq \mathrm{cl}(V)$. This gives $f((U - I) \cup (U - I)^\star) \subseteq \mathrm{cl}(V)$ and consequently $f(U^\star) \subseteq \mathrm{cl}(V)$. Since $\mathrm{cl}_{X^\star}(U \cap X^\star) = \mathrm{cl}(U \cap X^\star) = \mathrm{cl}(U \cap U^\star) \subseteq U^\star$, $f(\mathrm{cl}_{X^\star}(U \cap X^\star)) \subseteq \mathrm{cl}(V)$. Therefore $f/X^\star$ is θ-continuous. ■

From the remarks following Corollary 6.10 the following corollary (Lemma, [**14**]) is immediate.

COROLLARY 7.8. *Let* $f\colon (X, \tau) \to Y$ *be a function, with* $\mathscr{I}$ *an ideal on* X, *and* Y *regular. If* $f\colon (X, \tau^\star) \to Y$ *is continuous, then* $f/X^\star\colon (X^\star, \tau/X^\star) \to Y$ *is continuous.* ■

Our next result is in [**14**] for σ-ideals.

THEOREM 7.9. *Let* $f\colon (X, \tau) \to (Y, \sigma)$ *be a function, with* $\mathscr{I}$ *an ideal on* X, $X - X^\star \in \mathscr{I}$ *and* Y *regular. The following are equivalent*:

(a) $f\colon (X, \tau^\star) \to Y$ *is continuous,*

(b) $f/X^\star\colon (X^\star, \tau/X^\star) \to Y$ *is continuous, and*

(c) $f^{-1}(V) = U - I$ *with* $U \in \tau$ *and* $I \in \mathscr{I}$ *for every* $V \in \sigma$ (*i.e.*, $f\colon (X, \beta(\mathscr{I}, \tau)) \to (Y, \sigma)$ *is "continuous"*).

Proof. (a) → (b) by Corollary 7.8. It also follows by Corollary 6.9. To show (b) → (c) notice that

$$f^{-1}(F) = \left(f^{-1}(F) \cap X^\star\right) \cup \left(f^{-1}(F) \cap (X - X^\star)\right).$$

If F is closed in (Y, σ), then $f^{-1}(F) \cap X^\star = (f/X^\star)^{-1}(F)$ is closed in $(X^\star, \tau/X^\star)$ and consequently in (X, τ), and since $f^{-1}(F) \cap (X - X^\star) \in \mathscr{I}$, $f^{-1}(F) = P \cup I$ with P closed in (X, τ) and $I \in \mathscr{I}$. This is clearly equivalent to (c). Since (c) → (a) is obvious, the proof is complete. ■

Since the compatibility of τ and $\mathscr{I}$ implies $X - X^\star \in \mathscr{I}$, Theorem 7.9 is valid if "$X - X^\star \in \mathscr{I}$" is replaced by "$\tau \sim \mathscr{I}$." This then implies one of the main results in [**14**].

COROLLARY 7.10. *Let f: $(X, \tau) \to (Y, \sigma)$ be a function. with Y regular and $\mathscr{I}$ a σ-ideal of X. If either of the following holds*:

I. *(X, τ) is hereditarily Lindelöf, or*

II. *$\mathscr{I}$ is the σ-ideal of meager sets in (X, τ).*

Then the following are equivalent:

(a) *f/X_0 is continuous for some closed $X_0 \subseteq X$ with $X - X_0 \in \mathscr{I}$,*

(b) *$f^{-1}(V) = U - I$ with $U \in \tau$ and $I \in \mathscr{I}$ for every $V \in \sigma$.*

Proof. Note that in both cases, $\tau \sim \mathscr{I}$. ■

REFERENCES

1. P. Bankston, The total negation of a topological property, *Illinois J. of Math.*, 23 (1979) 241–252.
2. H. D. Block and B. Cargal, Arbitrary Mappings, *Proc. Amer. Math. Soc.*, 3 (1952) 937–941.
3. H. Blumberg, New properties of all real functions, *Trans. Amer. Math. Soc.*, 24 (1922) 113–128.
4. _____, Arbitrary point transformations, *Duke Math. J.*, 11 (1944) 671–685.
5. N. Bourbaki, General Topology, Addison-Wesley, Mass., 1966.
6. B. Cargal, Generalization of continuity, *Proc. Iowa Acad. Sci.*, 60 (1953) 477–481.
7. S. G. Crossley and S. K. Hildebrand, Semi-topological properties, *Fund. Math.*, 74 (1972) 233–254.
8. S. V. Fomin, Extensions of topological spaces, *Ann. of Math.*, (2) 44 (1943) 471–480.
9. G. Freud, Ein beitrag zu dem sätze von Cantor and Bendixson, *Acta Math. Hung.*, 9 (1958) 333–336.
10. H. Hashimoto, On some local properties on spaces, *Math. Japonica*, II (1952) 127–134.
11. _____, On the *topology and its application, *Fund. Math.*, 91 (1976) 5–10.
12. E. Hayashi, Topologies defined by local properties, *Math. Ann.*, 156 (1964) 205–215.
13. E. Hewitt, A problem in set theoretic topology, *Duke Math. J.*, 10 (1943) 309–333.
14. J. Kaniewski and Z. Piotrowski, Concerning continuity apart from a meager set, *Proc. Amer. Math. Soc.*, 98 (1986) 324–328.
15. M. R. Kirch, A class of space in which compact sets are finite, *Amer. Math. Monthly*, 76 (1969) 42.
16. K. Kuratowski, Topologie I, Warszawa, 1933.
17. _____, Topology, Vol. I, Academic Press, New York, 1966.
18. K. Kuratowski and W. Sierpinski, Le théorème de Borel-Lebesgue dans la théorie des Ensembles Abstraits, *Fund. Math.*, 2 (1921) 172–178.
19. N. Levine, On the equivalence of compactness and finiteness in topology, *Amer. Math. Monthly*, 75 (1968) 178–180.
20. N. F. G. Martin, Generalized condensation points, *Duke Math. J.*, 28 (1961) 507–514.
21. O. Njåstad, On some classes of nearly open sets, *Pacific J. Math.*, 15 (1965) 961–970.
22. _____, Remarks on topologies defined by local properties, *Avh. Norske Vid.-Akad. Oslo I(N.S)*, 8 (1966) 1–16.
23. J. C. Oxtoby, Measure and Category, Springer-Verlag, 1980.
24. _____, The kernel operation on subsets of a T_1-space, *Fund. Math.*, 90 (1976) 205–224.

25. I. L. Reilly and M. K. Vamanamurthy, On hereditarily Lindelöf spaces, *Bull. Austral. Math. Soc.*, 21 (1980) 357–362.
26. P. Samuels, A topology formed from a given topology and ideal, *J. London Math. Soc.*, (2) 10 (1975) 409–416.
27. S. Scheinberg, Topologies which generate a complete measure algebra, *Advances in Math.*, 7 (1971) 231–239.
28. P. L. Sharma, A class of spaces in which compact sets are finite, *Canad. Math. Bull.*, 24 (3) (1981) 373–375.
29. L. A. Steen and J. A. Seebach, Jr., Counterexamples in Topology, Springer-Verlag, New York, 1978.
30. R. Vaidyanathaswamy, Set Topology, Chelsea Publishing Company, 1960.
31. _____, The localization theory in set-topology, *Proc. Indian Acad. Sci.*, 20 (1945) 51–61.
32. A. Wilansky, Topology for Analysis, Ginn, Mass., 1970.

LETTERS TO THE EDITOR

Editor:

I have read the letter by Peter F. Zoll in Vol 96, number 8, p. 709, of this MONTHLY. I would like give a partial answer to Mr Zoll's questions. Regarding the alleged burning of Valmes in 1486 by Tomás de Torquemada because he had solved the general quartic equation, I have found no reference concerning either Valmes or his alleged punishment. However, this fact seems contradictory to known historical documentation, namely, that modern Algebra (Italian and German schools) was introduced into Spain circa 1552 by the German scholar Marc Aurel, [4], [7] p. 97. On the other hand, Zoll's letter gives the impression that there were no Spanish mathematicians to speak of in that period. It is this point I want to address in this letter.

Concerning the development of Algebra in Spain during the XVI century, one can come up with two names associated with highly original contributions, namely: Fray Juan de Ortega, whose approximation methods for square and cube roots have been commented on in [1], [4] and mainly by Rey Pastor in [7], who really shows a grasp and understanding of Ortega's method. The Portuguese mathematician Pedro Núñez's celebrated solution to the problem of the loxodromic curve made him famous in a world concerned with scientific navigation. Núñez was a professor at Salamanca, Spain, and after Aurel, was the author of a celebrated treatise on Algebra, written in Spanish, that contained many original views on the modern Algebra of the time. Other algebraists of the time that were mentioned in [4], [6], and [7] are J. Pérez de Moya and A. Rocha.

It is, however, in the field of infinitesimal Arithmetic and Mechanics that the Iberian mathematicians excelled during the XVI century. The two names that seem to be at the forefront of the Iberian school of Paris-Salamanca-Alcalá de Henares are Alvaro Thomás and Fray Domingo de Soto. A discussion of the achievements of this school can be consulted in [3], but are mainly in [8] and [9].

According to Pierre Duhem, [3] pp. 263–562, "Dominique de Soto et la escolastique parisienne," the science of the so-called Spanish school of Paris was the step that allowed the early Mechanics of the masters of Merton College (Oxford, circa 1350), Thomas Brawardine, Suiseth, etc., to mature into the views of Galileo and Galileo's Science. With reference to this particular point, see [2] chap. 11, [8] and [9].

Domingo de Soto, Dominican priest, Spanish mathematician, physicist and philosopher, virtuoso "calculator," educated in Paris and later professor at Salamanca, was the first scientist to describe the law of the falling bodies as a uniformly accelerated motion. His treatise "Super octo libros physicorum questiones," Salamanca 1555, earlier editions: 1551 and 1545, see Wallace [8] and [9], also M. Clagett [2] and Duhem [3], contains this result commonly attributed to Galileo. Pierre Duhem, [3], was the first to rectify this mistake. Domingo de Soto has been described as the central figure among the Parisian "calculatores" of the XVI century and the key link in the school whose output ultimately led to the Mechanics developed by Galileo, [3]. Concerning the diffusion of the works of the Spanish masters of Paris-Salamanca, Wallace points out in [8] p. 232 "More important, perhaps the works of the two Spanish Jesuits, Francisco Toledo and Benito Pereyro, the first of whom surely studied under Soto, were already widely diffused throughout the Italian peninsula. Spain's pre-eminence as world power during the sixteenth

century assured the diffusion of her books, as well as the exportation of her scholars. Galileo in his student notebooks mentions the works of Soto, Toledo and Pereyro, so these must have been known at least to his professors..."

Regarding the school of the Iberian mathematicians of Paris and Salamanca one can safely say: "They lay down the axiomatic and the formalization of the Calculus brought up by Brawardine and the masters of Merton College to study problems of motion." The infinitesimal character and essentially modern scope of this Calculus has been pointed out by many scholars, among them, H. Wieleitner, **[10]**, and Julio Rey Pastor [7]. The names associated with this school are: Martínez Silíceo, Pedro Ciruelo, Alvaro Thomás, Juan de Celaya, Luis Núñez Coronel and Diego Diest, to mention a few that achieved great fame. Perhaps the Portuguese Alvaro Thomás was the most conspicuous figure of this school. His academic life seems to have developed in Paris, where he wrote his treatise "Liber de triplici motu proportionibus annexis Magistri Alvari Thome Ulixbonensis philosophicas Suiseth calculationes ex parte declarans," Paris 1509. Among the novelties introduced by Alvaro Thomás we find a theory of series, which had very much the character of what we call today power series. The problems of convergence and divergence were clearly stated in the typical cases studied by Thomás. H. Wieleitner has shown in **[10]** Alvaro's mathematical originality and virtuosity in his approach to the theory of numerical series. Thomás was perhaps one of the first scholars of this school to state a general mean value theory in the context of theory of motion, (for the case of a motion more general than the uniformly accelerated one) **[3]**, **[10]**. Dr. Julio Rey Pastor has stated in [7] that Thomas deserves together with Oresme a position of honor in the history of Calculus.

I like to close this letter by commenting on D. Alonso de la Veracruz (the father of Mexican philosophy) who wrote "Physica speculatio," in 1557 in Mexico, intended to be used at the University of Mexico (founded in 1553). This treatise was very much inspired by the work of Soto and constitutes an early Hispanic American contribution to the development of Calculus and modern Mechanics, see **[8]** for further details.

[1] M. Cantor "Vorlesungen ueber die Geschichte der Mathematik," Leipzig, 1892.
[2] Marshall Clagett, "Science of Mechanics in the Middle Ages," The University of Wisconsin Press, Madison, 1959.
[3] Pierre Duhem, "Etude sur Leonard de Vinci," III, Paris 1913.
[4] G. Enestroem, "Quelques remarques sur l'histoire des mathématiques en Spagne au 16e siècle," Bibliotheca Mathematica, 2, 1984, 33–36.
[5] G. Enestroem, "Rey Pastor. Los mathemáticos españoles del siglo XVI (1913)." Rezension. Bibliotheca Mathematica, dritte Folge, vierzehnter Band, 85–90, Leipzig 1913–1914.
[6] J. Rey Pastor. "Los matemáticos españoles del siglo XVI." Discurso leído en la solemne apertura del curso académico de 1913 a 1914, Oviedo, 1913. Commented by G. Enestroem, see [5] above.
[7] J. Rey Pastor. "Los matemáticos españoles del siglo XVI," Toledo, 1926.
[8] William A. Wallace, O. P.. "The Calculatores in Early Sixteenth Century Physics." British Journal for the History of Science, 4, 15, 1969, 221–232.
[9] William A. Wallace, O. P. "The Enigma of Domingo de Soto: Uniformiter Difformis and Falling Bodies," Isis, 59, part 4, 1968, 384–402.
[10] H. Wieleitner "Zur Geschichte der unendlichen Reihen im christlichen Mittelalter." Bibliotheca Mathematica, dritte Folge, vierzehnter Band, 150–168, 1913–1914.

Dr. Calixto P. Calderón
Department of Mathematics
Statistics and Computer Science
University of Illinois
Chicago, Illinois 60680

Editor:

I am writing in response to the article "The effect of prior calculus experience on introductory college calculus" by M. B. Burton which appeared in the April issue of the *American Mathematical Monthly*. In that article students in the introductory calculus course at the University of New Hampshire were surveyed concerning their high school calculus experience by appending an informational question to their final college calculus exam.

It is a well-established principle in statistical science that randomized experiments should be used in determining cause and effect relationships. Otherwise, underlying confounding factors may be the cause of the perceived effect. In this survey it may well have been that students who chose to study calculus in high school had a propensity to do better in college calculus with or without their high school calculus experience. This may have been due to a variety of reasons including the possibility that better students come from high schools which offer calculus courses or that better students choose to take calculus in high school.

It is mentioned in the article that the students with high school calculus experience and the students without high school calculus experience had roughly the same quantitative SAT scores. This observation does not justify the conclusion that high school calculus experience has a positive effect on grades in college calculus. Even if quantitative SAT scores are a good predictor of performance in college calculus the remaining variation amongst students with equal SAT scores could again be due to some confounding factor such as those described above.

In summary, although it is possible that high school calculus experience has a positive effect on grades in college calculus, one cannot make such a conclusion based on a survey of this type.

Sincerely,

Tim Swartz
Dept of Math and Stats, SFU
Burnaby, B.C. Canada V5A1S6

Editor:

I have been wondering if you and the membership are prepared for the coming change.

The change I am referring to is a change in name. Our Association will soon change its name to Math Association of America. Some of us will soon be members of the American Math Society, some to the Society for Industrial and Applied Math. That's *Math*, not Mathematics.

Do you think I exaggerate? Recently an article appeared in the Boston Globe about mathematics which did not use the word "mathematics" once, just "math." A book publisher's advertisement in this MONTHLY recently referred to the American Math Society. Look carefully at a recent issue of *UME Trends*. Listen to your students. Probably you can all supply many examples.

Do you find this disturbing? Alarming? Do you care?

Anyone interested in joining *MATHEMATICIANS AGAINST MATH*?

R. E. Worth
Georgia State University
Atlanta, Georgia 30303

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

A Guessing Game of Bill Sands, and Bernardo Recamán's Barranca

RICHARD K. GUY

Bill Sands devised a neat little problem which was used in the Alberta High School Mathematics Scholarship Examination, 87-02-10 [**4**].

I think of a number from 1 to 12, and you are allowed four guesses, to each of which I shall give you the correct one of the following three responses:

(a) you're exactly right,

(b) you're near (meaning "within 3"),

(c) you're far (otherwise).

The last guess must give the right answer: it's not sufficient to be able to say, "Yes, now I know the answer."

At first glance this looks easy: with three outcomes for each of four guesses, the number of responses, 3^4, is many times more than the number of answers, 12. But the responses are very interdependent, and only a few can occur. We'll wait till the end before giving an answer, in case you want to try it out for yourself.

It's reminiscent of the famous counterfeit coin problem (12 pennies with one dud, see [**1**, **3**], for example) but here you can't find an "unconditional" solution: later guesses have to depend on feedback from earlier ones. However, for the pennies it's possible to state ahead of time just which coins will be placed in each pan of the balance at each weighing.

The reason for mentioning the problem here is to ask for the general result:

How many guesses are needed to determine a number chosen from $1, 2, \ldots, N$ if you are told that each guess is either correct, or within k, or not?

If $k \geqslant N - 1$, then N guesses may be needed if your guesses are constrained to the numbers $1, 2, \ldots, N$, but if not, or in any case if $(N - 1)/2 \leqslant k \leqslant N - 2$, then some economy is possible, e.g., the responses to an initial guess of 1 partition the population into the sets $\{1\}, \{2, \ldots, k + 1\}, \{k + 2, \ldots, N\}$ and it might seem that $\max\{k, N - k - 1\}$ further guesses could be needed. But if this max is k, i.e., if $k \geqslant (N - 1)/2$, further economy follows from the wrong guess of N: the response indicates which of the two sets $\{2, \ldots, N - k - 1\}$ and $\{N - k, \ldots, k + 1\}$ contains the number. When the cardinality of the surviving candidates is small compared with k (as it must eventually become) it may still pay to "waste" a guess in order to halve the remaining possibilities. The problem is to quantify this.

If this problem isn't sufficiently general, then answer the problem with four or more possible responses, e.g., with "within k_1, or within k_2" in place of "within k." And after Ivan Niven's delightful article [**2**] we can ask the variants:

What if the responder lies exactly once? At most once?

Bernardo Recamán, while visiting schools in Colombia, has been asked more than once for a game which could stimulate children to practice their multiplication tables, so he invented the game of Barranca. The name is short for Barrancabermeja, a town on the Magdalena River in central Columbia; it also means "deep hollow," "ravine," or "precipice." Barranca is played with sixteen cards, numbered $1, 2, \ldots, 16$. Two players alternately choose a card, until each has eight. The winner is the one who has a (sub)set of numbers whose product is 220, or, if neither player has that product, then the winner is the one who can display a product nearest to 220. It is possible for ties to occur, but it seems likely that, with best play, the first player has a winning strategy. Is there a "strategy stealing" argument which proves this? Recamán doesn't know of any simple strategy.

Can the first player win? If so, how? What happens if there are three (or more) players? Are there better numbers to use in place of 16 and 220?

Here's a solution to the Bill Sands problem: make the first two guesses 6 and 8. After the responses "near, near," "near, far," "far, near," and "far, far," guess respectively 5 (or 9), 3 (or 4), 10 (or 11), and 1 (or 2).

REFERENCES

1. Freeman J. Dyson, Note 1931. The problem of the pennies, *Math. Gaz.* 30 (1946) 231–234.
2. Ivan Niven, Coding theory applied to a problem of Ulam, *Math. Mag.*, 61 (1988) 275–281.
3. Cedric A. B. Smith, The counterfeit coin problem, *Math. Gaz.*, 31 (1947) 31–39.
4. Robert E. Woodrow, The Olympiad Corner, *Crux Mathematicorum*, 13 #2 (Feb. 1987) 38, Problem 4; 14 #6 (Jun. 1988) 170–171.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

Polynomials With Large Numbers of Prime Values

BETTY GARRISON
Department of Mathematical Sciences, San Diego State University, San Diego, CA 92182

Let $f(x)$ denote a polynomial with integer coefficients, and let $G(f)$ be the largest integer G such that G divides $|f(x)|$ for all integers x. Bouniakowsky conjectured [**1**] in 1857 that if f is nonconstant and $G(f) = 1$, then there exist infinitely many integers x such that $|f(x)|$ is prime. The case where f has degree one is known today as Dirichlet's theorem on primes in arithmetic progressions. At present it remains unknown whether there exists even a single polynomial of degree greater than one which is prime for infinitely many integer values x, but evidence is accumulating in the literature to indicate that this is very likely to be the case. Sierpinski [**4**] has shown that for every natural number m there exists a natural number n such that $x^2 + n$ is prime for more than m distinct natural numbers x. Iwaniec [**3**] proved that if $g(x) = ax^2 + bx + c$ is an irreducible polynomial with integer coefficients such that $a > 0$ and $c \equiv 1 \pmod 2$ then there exist infinitely many integers x such that $g(x)$ is either prime or a product of two primes. Others have obtained results similar to Iwaniec's (but a little weaker) for polynomials of any degree $\geqslant 2$.

In this note we first extend Sierpinski's result to polynomials of degree $k \geqslant 2$ and then to polynomials $x^2 + a^2$, where a is a natural number.

THEOREM 1. *Let $k \geqslant 2$ be an integer and let M be an arbitrarily large real number. Then there exist positive integers c, d such that $x^k + c$, likewise $x^k - d$, is prime for more than M positive integers x.*

Proof. Let $[x]$ denote the largest integer $\leqslant x$, and let $\pi(x)$ denote the number of primes in the interval $[1, x]$. From the prime number theorem, $\pi(x) \sim x/\ln x$, and it follows that for every ε in the interval $(0, 1)$ there exists $y(\varepsilon)$ such that $\pi(x^k) > (1 - \varepsilon)x^k/k \ln x$ for every $x^k > y(\varepsilon)$. The proof of the theorem will depend on showing that for each pair k, M there exists an interval $[1, J]$ containing more than MJ representations $p - n_p^k$ for distinct primes p and associated positive integers n_p. This will imply that there is at least one integer c in $[1, J]$ equal to more than M of these representations. In order to accomplish this, choose a fixed ε in $(0, 1)$ and a positive integer N such that (1) $N^k > y(\varepsilon)$, (2) $(1 - \varepsilon)N/k(k + 1)\ln N > M$, and (3) $N > 2^k - k$. Then for every prime p in $[1, N^k]$ there exists a positive integer $n_p = [p^{1/k}] < N$ such that $n_p^k < p < (n_p + 1)^k$. Hence

$$p - n_p^k < (n_p + 1)^k - n_p^k < kn_p^{k-1} + n_p^{k-2}\left\{\binom{k}{2} + \binom{k}{3} + \cdots + 1\right\}$$

$$< kn_p^{k-1} + n_p^{k-2}(2^k - k) < (k + 1)N^{k-1}.$$

Thus all the integers $p - n_p^k$ are in the interval $[1, (k + 1)N^{k-1}]$, and since there are $\pi(N^k)$ such values, there exists an integer c in the interval $[1, (k + 1)N^{k-1}]$ such

that $p - n_p^k = c$ for at least $((1 - \varepsilon)N^k/k \ln N)/(k + 1)N^{k-1} > M$ primes p. The conclusion is reached for $x^k - d$ by repeating the argument above with $p - n_p^k$ replaced by $(n_p + 1)^k - p$.

THEOREM 2. *Let M be an arbitrarily large real number. Then there exists a positive integer c such that $x^2 + c^2$ is prime for more than M positive integers x.*

Proof. Let $\pi(x; 4, 1)$ denote the number of primes $p \equiv 1 \pmod 4$ in the interval $[1, x]$, and recall that the prime number theorem for primes in arithmetic progressions gives $\pi(x; 4, 1) \sim x/2 \ln x$. Thus for a fixed ε in the interval $(0, 1)$ we can choose a positive integer N sufficiently large so that $\pi(N^2; 4, 1) > (1 - \varepsilon)N^2/4 \ln N$, and $(1 - \varepsilon)N/2 \ln N > M$. It is well known that for each prime $p \equiv 1 \pmod 4$ there exists a unique pair a_p, b_p of positive integers such that a_p is even and $a_p^2 + b_p^2 = p$. (See [2], pp. 264–265, for example.) Since there are at most $N/2$ even integer squares in $[1, N^2]$, it follows that there exists an even integer c such that $a_p = c$ for at least $\pi(N^2; 4, 1)/(N/2) > M$ primes p. For these primes p we have $c^2 + b_p^2$ prime.

REFERENCES

1. V. Bouniakowsky, Sur les diviseurs numeriques invariables des fonctions rationelles entieres, *Mem. Acad. Sci. St. Petersburg*, 6 (1857) 305–329.
2. D. M. Burton, Elementary Number Theory, Allyn and Bacon, 1976.
3. H. Iwaniec, Almost-primes represented by quadratic polynomials, *Inventiones Mathematicae*, 47 (1978) 171–188.
4. W. Sierpinski, Les binômes $x^2 + n$ et les nombres premiers, *Bull. Soc. Royale Sciences Liège*, 33 (1964) 259–260.

On Inequivalent Group Extensions

F. B. CANNONITO
Department of Mathematics, University of California, Irvine, CA 92717

The purpose of this note is to explicitly construct inequivalent group extensions which are not isomorphisms of extensions. We review the basics and terminology.

Let $N \overset{\mu}{\rightarrowtail} E \overset{\varepsilon}{\twoheadrightarrow} G$ be an extension of groups with N abelian, so that $G \simeq E/N$. If $\overline{N} \overset{\bar{\mu}}{\rightarrowtail} \overline{E} \overset{\bar{\varepsilon}}{\twoheadrightarrow} \overline{G}$ is another extension, then a *morphism* of extensions is a triple (α, β, ν) of group homomorphisms such that following diagram (1) commutes

$$\begin{array}{ccccc} N & \overset{\mu}{\rightarrowtail} & E & \overset{\varepsilon}{\twoheadrightarrow} & G \\ {\scriptstyle\alpha}\downarrow & & {\scriptstyle\beta}\downarrow & & {\scriptstyle\nu}\downarrow \\ \overline{N} & \overset{\bar{\mu}}{\rightarrowtail} & \overline{E} & \overset{\bar{\varepsilon}}{\twoheadrightarrow} & \overline{G}. \end{array} \tag{1}$$

The triple $(0, 0, 0)$ will be called the *trivial morphism*. A morphism of extensions of the form $(1, \beta, 1)$ between $N \rightarrowtail E \twoheadrightarrow G$ and $N \rightarrowtail \overline{E} \twoheadrightarrow G$ is termed an equivalence. A more general type of morphism is an *isomorphism*; this is a morphism of the form $(1, \beta, \nu)$ (resp., $(\alpha, \beta, 1)$) from $N \rightarrowtail E \twoheadrightarrow G$ to $N \rightarrowtail \overline{E} \twoheadrightarrow \overline{G}$ (resp., $N \rightarrowtail E \twoheadrightarrow G$ to $\overline{N} \rightarrowtail \overline{E} \twoheadrightarrow G$) where α or ν is an isomorphism of groups. It follows from the five

lemmas that β is also an isomorphism of groups. Equivalence and isomorphism of extensions are equivalence relations. Denoting by $\mathrm{Ext}(G, N)$ the set of equivalence classes of abelian extensions of N by G, it is well known that $\mathrm{Ext}(G, N)$ can be made into an abelian group. (See, for example, Rotman [**1** Chap. 10].)

As an example of this development, the student is led to a study of $\mathrm{Ext}(Z_p, Z_p)$, p prime. Indeed, $\mathrm{Ext}(Z_p, Z_p) \simeq Z_p$ and so $\mathrm{Ext}(Z_p, Z_p)$ has p elements. But if $Z_p \rightarrowtail E \twoheadrightarrow Z_p$ is exact, then E has order p^2 so that $E \simeq Z_{p^2}$ or $E \simeq Z_p \oplus Z_p$. This shows for p odd that there exist inequivalent extensions of N by G with isomorphic middle terms. Specializing to $p = 3$ it can be easily seen that given any morphism (α, β, ν) of inequivalent extensions of Z_3 by Z_3, where β is an isomorphism, one must have $\alpha = 1$ or $\nu = 1$ but, of course, not both; so any morphism here is necessarily an isomorphism.

We will now exhibit a morphism of inequivalent extensions of Z_p by Z_p for $p > 3$, which is not an isomorphism. First, we construct inequivalent extensions.

Consider the following extensions

$$\begin{array}{ccccc} E_1\colon & Z_p & \overset{\mu}{\rightarrowtail} & Z_{p^2} & \overset{\varepsilon}{\twoheadrightarrow} Z_p \\ & & & \downarrow \beta & \qquad p > 2, \\ E_2\colon & Z_p & \overset{\mu}{\rightarrowtail} & Z_{p^2} & \overset{-\epsilon}{\twoheadrightarrow} Z_p \end{array} \tag{2}$$

where μ is $x \mapsto px$ and ε is $y \mapsto y + Z_p$ and β is any group isomorphism.

LEMMA 1. *E_1 is not equivalent to E_2.*

Proof. Suppose the diagram (2) could be completed so that it commutes with respect to the morphism $(1, \beta, 1)$. Then $1 + Z_p = -\beta(1) + Z_p$ so that (a) $\beta(1) \equiv -1 \pmod{p}$. Furthermore, $p \equiv p\beta(1) \pmod{p^2}$ and thus (b) $\beta(1) \equiv 1 \pmod{p}$. Hence $2 \equiv 0 \pmod{p}$ which is impossible for $p > 2$.

We see from Lemma 1 that the three elements in $\mathrm{Ext}(Z_3, Z_3)$ are the equivalence classes of the extension $Z_3 \oplus Z_3$ and the two inequivalent extensions of the form Z_9.

LEMMA 2. *If $p > 3$ then the map $\beta\colon x \to 2x$ is an isomorphism $Z_{p^2} \to Z_{p^2}$ such that there exists a morphism (α, β, ν) of diagram (2) with neither $\alpha = 1$ nor $\nu = 1$.*

Proof. Using commutativity on the right we see that if $\nu \neq 1$ then, (writing $\nu(x + Z_p)$ as $\nu x + Z_p$) (*a*) $\nu \not\equiv 1 \pmod{p}$ and $\nu x \equiv -2x \pmod{p}$ and so for $x = 1$ $\nu \equiv -2 \pmod{p}$. Using commutativity on the left, we have $p\alpha(1) \equiv p\beta(1) \pmod{p^2}$ or $p\alpha \equiv p\beta \pmod{p^2}$ or $\alpha \equiv \beta \pmod{p}$. Since $\alpha \neq 1$, $\alpha \not\equiv 1 \pmod{p}$. But β is $x \mapsto 2x$, so for $x = 1$ we get the system (†) $\alpha \equiv 2 \pmod{p}$ and $\alpha \not\equiv 1 \pmod{p}$. Taking $p > 3$ we see both $\beta(1) \equiv 1 \pmod{p}$ and the system (†) have solutions for α and β and clearly for such solutions $x \mapsto \alpha x$ and $y \mapsto \beta y$ are isomorphisms which make the diagram commute. Thus, (α, β, ν) is a morphism with $\alpha \neq 1$ and $\nu \neq 1$.

We note for $p = 3$ there is no solution of $\nu \not\equiv 1 \pmod{3}$ and $\nu \equiv -2 \pmod{3}$ and so each nontrivial morphism of inequivalent extensions of Z_3 by Z_3 must be an isomorphism.

Hence we have proved the

THEOREM. *If p is a prime greater than 3 there exist nontrivial morphisms of inequivalent extensions of Z_p by Z_p which are not isomorphisms.*

The author wishes to thank W. Messing for discussions which led to this note.

REFERENCE

1. J. J. Rotman, The Theory of groups, An Introduction, 3rd ed., Allyn and Bacon.

A Rearrangement Inequality and the Permutahedron

A. VINCE
Department of Mathematics, University of Florida, Gainesville, FL 32611

One chapter of the classic book "*Inequalities*" by Hardy, Littlewood, and Pólya [3] is dedicated to inequalities involving sequences with terms rearranged. The main example in that chapter is the following. Let $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ and $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$ be sequences of real numbers and π any permutation of the set $\{1, 2, \ldots, n\}$. Then

$$\sum_{i=1}^{n} a_i b_{n-i+1} \leqslant \sum_{i=1}^{n} a_i b_{\pi i} \leqslant \sum_{i=1}^{n} a_i b_i. \tag{1}$$

Hardy, Littlewood, and Pólya interpret the a_i as fixed distances along a rod and the b_i as weights to be suspended at these distances: To get the maximum moment with respect to an end of the rod hang the heaviest weights furthest from that end; to get the minimum moment hang the heaviest weights closest.

Many variations and generalizations of this rearrangement inequality exist. Three appear below, more at the end of this note, and Marshall and Olkin [5] contains a relatively recent survey. In this note *all sequences* $\{a_i\}$ *and* $\{b_i\}$ *are increasing* in the sense $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ and $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$, and *all sums and products are from* $i = 1$ *to* $i = n$ unless otherwise stated. Also $N = \{1, 2, \ldots, n\}$ and S_n denotes the set of all permutations of N.

Example 1. [7] For sequences with positive terms and for all $\pi \in S_n$

$$\prod(a_i + b_{n-i+1}) \geqslant \prod(a_i + b_{\pi i}) \geqslant \prod(a_i + b_i).$$

Example 2. [8] For $p \geqslant 1$ and all $\pi \in S_n$

$$\sum|a_i - b_i|^p \leqslant \sum|a_i - b_{\pi i}|^p \leqslant \sum|a_{n-i+1} - b_i|^p.$$

Example 3. [4] One generalization of the Hardy-Littlewood-Pólya inequality (1) is as follows. If the sequences have positive terms and f is an increasing convex function then for all $\pi \in S_n$

$$\sum f(a_i b_{n-i+1}) \leqslant \sum f(a_i b_{\pi i}) \leqslant \sum f(a_i b_i).$$

Recall that f is *increasing* if $f(x) \geqslant f(y)$ whenever $x \geqslant y$, and f is *convex* if

$f(\alpha x + (1 - \alpha)y) \leqslant \alpha f(x) + (1 - \alpha)f(y)$ for all $0 \leqslant \alpha \leqslant 1$. In the differentiable case, of course, convex is equivalent to $f''(x) \geqslant 0$.

The purpose of this note is to show how the permutahedron leads to a very simple generalization of the Hardy-Littlewood-Pólya inequality, from which the inequalities above and many other rearrangement inequalities immediately follow. In [5] rearrangement inequalities are derived using *majorization*, which is a partial order on the set of vectors in n-dimensional Euclidean space. This elegant method goes back at least to Schur [**10**] and is a unifying principal for many types of inequalities. The intention in this note is to use a much simpler partial order on S_n to obtain the rearrangement inequalities.

Let $\mathcal{I}$ denote the identity permutation $\mathcal{I}(i) = i$ of N and $\mathcal{I}_\star$ the reverse permutation $\mathcal{I}_\star(i) = n - i + 1, i = 1, \ldots, n$. An inversion of a permutation π of N is a pair $(\pi j, \pi k)$ such that $j < k$ and $\pi j > \pi k$. For example, (5, 3) is one of the four inversions of $\pi = 25134$. Now consider the directed graph P_n whose vertex set is S_n, and there is an edge (σ, π) directed from σ to π whenever vertex σ is obtained from vertex π by interchanging the elements of an inversion of the form $(\pi j, \pi(j + 1))$. P_n is sometimes called a permutahedron [**1**] and an example is shown in Figure 1. The transitive closure of P_n induces a partial order on the set S_n. Recall that the transitive closure is the "smallest" directed graph with the same vertex set as P_n and with the property that if (π, τ) and (τ, σ) are directed edges then so is (π, σ); the partial order on S_n is defined by $\sigma > \pi$ if there is an edge directed from σ to π in the transitive closure of P_n. A real valued function g: $S_n \to R$ is called *order preserving* if $g(\sigma) > g(\pi)$ whenever $\sigma > \pi$. This is all that is needed to prove the following theorem.

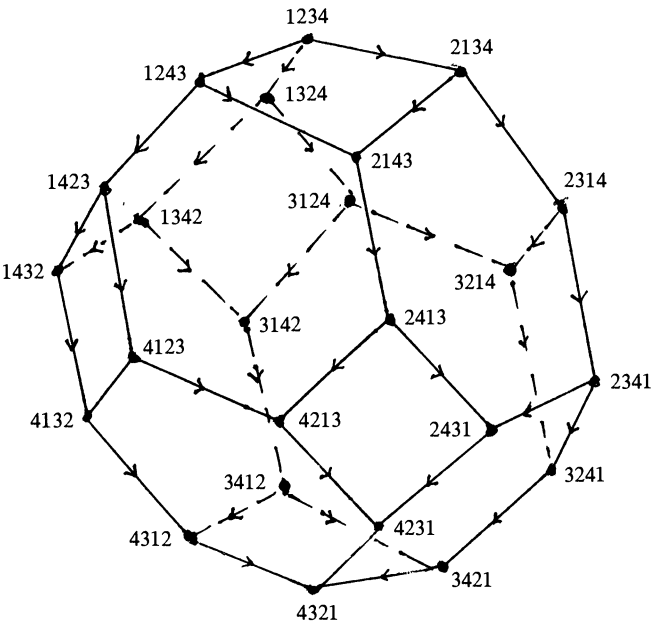


FIG. 1. Permutahedron P_4.

THEOREM. *Let $g_1, \ldots, g_n$ be real valued functions defined on an interval I. Then*

$$\sum g_i(b_{n-i+1}) \leqslant \sum g_i(b_{\pi i}) \leqslant \sum g_i(b_i) \tag{2}$$

for all sequences $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$ in I and all $\pi \in S_n$ if and only if

$$g_{i+1} - g_i \text{ is increasing on } I, \quad 1 \leqslant i < n. \tag{3}$$

Remark. If the functions g_i are differentiable, then it is clear that

$$g_1'(x) \leqslant g_2'(x) \leqslant \cdots \leqslant g_n'(x) \quad \text{for all } x \in I \tag{3'}$$

is equivalent to condition (3).

Proof. ($\Leftarrow$) Fix a sequence $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$ and let g: $S_n \to R$ be defined by $g(\pi) = \sum g_i(b_{\pi i})$. To show that g is order preserving it suffices to show that $g(\sigma) > g(\pi)$ whenever (σ, π) is an edge of the permutahedron:

$$\begin{aligned} g(\sigma) - g(\pi) &= \left[g_j(b_{\sigma j}) + g_{j+1}(b_{\sigma(j+1)})\right] - \left[g_j(b_{\pi j}) + g_{j+1}(b_{\pi(j+1)})\right] \\ &= (g_{j+1} - g_j)(b_{\pi j}) - (g_{j+1} - g_j)(b_{\pi(j+1)}) \geqslant 0. \end{aligned}$$

The last inequality follows from $b_{\pi j} < b_{\pi(j+1)}$ and the assumption that $g_{j+1} - g_j$ is increasing. Since $\mathscr{I}_\star \leqslant \pi \leqslant \mathscr{I}$ for any permutation π, also $g(\mathscr{I}_\star) \leqslant g(\pi) \leqslant g(\mathscr{I})$, which is precisely inequality (2).

($\Rightarrow$) By way of contradiction assume that $g_{m+1} - g_m$ is not increasing for some m. Then there exists $x > y$ such that $(g_{m+1} - g_m)(x) < (g_{m+1} - g_m)(y)$. Now choose any sequence $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_n$ in I with $b_m = x$ and $b_{m+1} = y$. Let π be the transposition $(m\ m+1)$. Then $\sum g_i(b_{\pi i}) =$

$$\begin{aligned} &\sum_{i \neq m,\, m+1} g_i(b_i) + g_m(x) + g_{m+1}(y) \\ &\quad > \sum_{i \neq m,\, m+1} g_i(b_i) + g_m(y) + g_{m+1}(x) = \sum g_i(b_i), \end{aligned}$$

contradicting inequality (2). □

The inequalities of Examples 1, 2, and 3, as well as those below, result by simply substituting the appropriate g_i in the theorem. For example, choosing $g_i(x) = a_i x$ yields the classic Hardy-Littlewood-Pólya inequality (1). Choosing $g_i(x) = -\log(a_i + x)$ yields Example 1; and choosing $g_i(x) = f(a_i x)$ gives Example 3. In each case it is an exercise to show that the g_i satisfy condition (3) or (3') in the theorem. For the Hardy-Littlewood-Pólya inequality condition (3) is immediate. For Example 3 the verification of condition (3) is a little tricky, but still elementary.

Example 4. Let $g_i(x) = f(a_i - x)$. If f is convex then for all $\pi \in S_n$

$$\sum f(a_i - b_i) \leqslant \sum f(a_i - b_{\pi i}) \leqslant \sum f(a_{n-i+1} - b_i).$$

With $f(x) = |x|^p$ this is Example 2.

Example 5. [2] Take $g_i(x) = g(a_i, x)$ where g is a real valued function of two variables defined on a domain $D = [a, b] \times [c, d]$. If

$$g(x_1, y_1) - g(x_1, y_2) - g(x_2, y_1) + g(x_2, y_2) \geqslant 0 \tag{4}$$

for all $x_1 \geqslant x_2$, $y_1 \geqslant y_2$ in the domain then

$$\sum g(a_1, b_{n-i+1}) \leqslant \sum g(a_i, b_{\pi i}) \leqslant \sum g(a_i, b_i) \tag{5}$$

for all $\pi \in S_n$. Note that for a function with continuous second derivatives condition (4) can be replaced by

$$\frac{\partial^2 g(x, y)}{\partial x \partial y} \geqslant 0 \quad \text{for all } (x, y) \in D. \tag{4'}$$

Example 6. Many inequalities can be generalized to more than two sequences. Let $\{a_i^1\}, \{a_i^2\}, \ldots, \{a_i^m\}$, $1 \leqslant i \leqslant n$, be *not necessarily increasing* real sequences and let $a_{(1)}, a_{(2)}, \ldots, a_{(n)}$ denote the sequence $a_1, a_2, \ldots, a_n$ in increasing order. Suppose $g(x_1, \ldots, x_m)$ satisfies condition (4) or (4') for every pair of variables. Then

$$\sum g\left(a_i^1, a_i^2, \ldots, a_i^m\right) \leqslant \sum g\left(a_{(i)}^1, a_{(i)}^2, \ldots, a_{(i)}^m\right). \tag{6}$$

This inequality follows directly by induction using the right inequality of (5) as the first step. Choosing $g(x_1, \ldots, x_m) = x_1 x_2 \cdots x_m$ and $g(x_1, \ldots, x_m) = -\log(x_1 + \cdots + x_m)$, respectively, in (6) results, for sequences of positive terms, in inequalities analogous to (1) and Example 2 [**9**]:

$$\sum a_i^1 a_i^2 \ldots a_i^m \leqslant \sum a_{(i)}^1 a_{(i)}^2 \cdots a_{(i)}^m$$

and

$$\prod\left(a_i^1 + a_i^2 + \cdots + a_i^m\right) \geqslant \prod\left(a_{(i)}^1 + a_{(i)}^2 + \cdots + a_{(i)}^m\right).$$

Example 7. Choose $g(x_1, \ldots, x_m) = f[\min(x_1, \ldots, x_m)]$ in (6). If f is an increasing function then

$$\sum f\left(\min_j a_i^j\right) \leqslant \sum f\left(\min_j a_{(i)}^j\right).$$

Taking $f(x) = x$ and $f(x) = \log(x)$, respectively, leads to inequalities of Minc [**6**]:

$$\sum \min_j a_i^j \leqslant \sum \min_j a_{(i)}^j$$

and

$$\prod \min_j a_i^j \leqslant \prod \min_j a_{(i)}^j.$$

Similar inequalities hold for the max function.

REFERENCES

1. C. Berge, Principles of Combinatorics, Academic Press, New York and London, 1971.
2. P. W. Day, Rearrangement inequalities, *Canad. J. Math*, 24 (1972) 930–943.
3. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, London, 1952.
4. D. London, Rearrangement inequalities involving convex functions, *Pacific J. Math.*, 34 (1970) 749–753.
5. A. W. Marshall and I. Olkin, Inequalities: Theory of Majorization and Its Applications, Academic Press, New York, 1979.
6. H. Minc, Rearrangements, *Trans. Amer. Math. Soc.*, 159 (1971) 497–504.

7. A. Oppenheim, Inequalities connected with definite Hermitian forms II, *Amer. Math. Monthly*. 61 (1954) 263–266.
8. Y. Rinott, Multivariate majorization and rearrangement inequalities with some applications to probability and statistics, *Israel J. Math.*, 15 (1973) 60–77.
9. H. D. Ruderman, Two new inequalities, *Amer. Math. Monthly*, 59 (1952) 29–32.
10. I. Schur, Uber eine Klasse von Mittelbildungen mit Anwendungen die Determinanten, Theorie Sitzungsber., *Berlin. Math. Gesellschaft*, 22 (1923) 9–20.

Musings on the Prime Divisors of Arithmetic Sequences

PATRICK MORTON
Department of Mathematics, Wellesley College, Wellesley, MA 02181

Of the early proofs one usually sees in a number theory course, the most beautiful is the proof, due to Euclid, that there are infinitely many primes. This theorem may be formulated as follows.

If $\{a_n\}_{n=1}^{\infty}$ is any sequence of integers, and p is a prime for which $p|a_n$ for some n, p is called a *prime divisor* of the sequence $\{a_n\}_{n=1}^{\infty}$. (See **[2]**.) Euclid's theorem says that the sequence $\{n\}_{n=1}^{\infty}$ has an infinite set of prime divisors. What other sequences have this property?

For example, consider the sequences whose terms are defined by the following formulas:

1) $a_n = f(n)$, where $f \in \mathbf{Z}[x]$ is a nonconstant polynomial;
2) $b_n = [\pi n^2]$, where brackets denote the greatest integer;
3) $c_n = [\pi n^2]^2 - [\pi n^2]\left[\frac{\sqrt{2}}{en}\right] + \left[\frac{\sqrt{2}}{en}\right]^2$;
4) $d_n = 2^n + 1$.

Which of these sequences has an infinite number of prime divisors?

The answer is, of course, that they all do. The fact that $\{a_n\}_{n=1}^{\infty}$ does was first proved in an elementary way by Schur **[8]**, and is usually stated as follows. (See **[2]** and **[3]** for other proofs and more on the prime divisors of polynomials.)

THEOREM 1. *Let $f(x) \in \mathbf{Z}[x]$ be nonconstant. Then the congruence*

$$f(x) \equiv 0 (\text{mod } p)$$

has a solution $x \in \mathbf{Z}$ for infinitely many primes p. In other words, infinitely many primes divide the terms of the sequence $\{f(n)\}_{n=1}^{\infty}$.

The purpose of this note is to give a surprising proof of this result using a well-chosen infinite series, and then to see where the proof leads. It will turn out that the proof can be generalized to show that the sequences $\{b_n\}_{n=1}^{\infty}$ and $\{c_n\}_{n=1}^{\infty}$ both have infinite sets of prime divisors, but that the same proof cannot decide this question for the simpler sequence $\{d_n\}_{n=1}^{\infty}$! The proof will also be a surprise in that it hides an algebraic structure.

1. A proof of Schur's theorem. Assume theorem 1 is false for some non-constant $f(x) \in \mathbf{Z}[x]$, and let $m = \deg f$. Then for $n \in \mathbf{Z}$, $f(n) = 0$ or (by the fundamental theorem of arithmetic)

$$f(n) = \pm p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k},$$

for a fixed set $\{p_1, p_2, \ldots, p_k\}$ of primes and exponents $a_i \geq 0$ in $\mathbf{Z}$. Furthermore, since f is a polynomial, f can take on a given value at most m times. Choose N large enough so that $f(n) \neq 0$ for $n \geq N$ and form the infinite series

$$S = \sum_{n=N}^{\infty} \frac{1}{|f(n)|^{1/m}}. \tag{1}$$

By the above remarks $|f(n)|^{1/m} = p_1^{a_1/m} p_2^{a_2/m} \cdots p_k^{a_k/m}$, for some $a_1, a_2, \ldots, a_k$, and this holds, for a given set of integers $a_1, a_2, \ldots, a_k$, for at most $2m$ values of n. Hence we have the estimate

$$\begin{aligned} S &\leq \sum_{a_1, \ldots, a_k = 0}^{\infty} \frac{2m}{p_1^{a_1/m} \cdots p_k^{a_k/m}} \\ &= 2m \prod_{r=1}^{k} \sum_{a_r=0}^{\infty} p_r^{-a_r/m} \qquad (2) \\ &= 2m \prod_{r=1}^{k} \frac{1}{1 - p_r^{-1/m}} < \infty, \end{aligned}$$

so S is a convergent series. But this is impossible, since

$$|f(n)|^{1/m} \leq an,$$

for some constant $a > 0$, so that S diverges like the harmonic series. This contradiction proves theorem 1.

Let us now see what paths this proof can lead us down. Without any changes the same proof yields the following stronger theorem.

THEOREM 2. *Let f: $\mathbf{N} \to \mathbf{Z}$ be any arithmetic function which satisfies*
(a) $|f(n)| \leq an^m$, *for sufficiently large n and some a, $m > 0$;*
(b) *the number of solutions of $f(n) = k$ is bounded, for all $k \in \mathbf{Z}$.*
Then the sequence $\{f(n)\}_{n=1}^{\infty}$ has an infinite number of prime divisors.

As a simple (nonpolynomial) example of a function satisfying the hypotheses of theorem 2 take

$$f(n) = [\theta n^{\alpha}],$$

where θ and α are positive real numbers and $\alpha \geq 1$. Property (a) is clear, and property (b) is not hard to show. For example, if $f(n) = k$, then

$$k \leq \theta n^{\alpha} < k + 1,$$

or

$$\left(\frac{k}{\theta}\right)^{1/\alpha} \leq n < \left(\frac{k+1}{\theta}\right)^{1/\alpha}$$

The number $N(k)$ of integers k, for which the last inequality holds, satisfies

$$\begin{aligned} N(k) &\leq \left(\frac{k+1}{\theta}\right)^{1/\alpha} - \left(\frac{k}{\theta}\right)^{1/\alpha} + 1 \\ &= \frac{1}{\theta\alpha}\left(\frac{c}{\theta}\right)^{(1/\alpha)-1} + 1, \end{aligned}$$

for some c with $k < c < k + 1$, by the mean value theorem applied to $(x/\theta)^{1/\alpha}$. Since $\alpha \geq 1$ and $c > k$, this obviously implies $N(k)$ is bounded, and proves (b).

For example, each of the sequences

$$\{[\pi n^2]\}_{n=1}^{\infty} \quad \text{and} \quad \left\{\left[\frac{\sqrt{2}}{en}\right]\right\}_{n=1}^{\infty}$$

has an infinite set of prime divisors.

2. The proof hides algebraic structure. But now consider the conditions in theorem 2. If f, g both satisfy (a) and (b), and g: $\mathbf{N} \to \mathbf{N}$, then the composition sequence $\{f(g(n))\}_{n=1}^{\infty}$ satisfies (a) and (b) also!

This is quite easy to see. To prove (a): if $|f(n)| \leq an^m$ and $|g(n)| \leq bn^{m'}$, then

$$|f(g(n))| \leq a(g(n))^m \leq ab^m n^{mm'}.$$

For the proof of (b): if $f(n) = k$ has at most K solutions for any k and $g(m) = j$ has at most J solutions for any j, then $f(g(n)) = k$ has at most KJ solutions for any k. This gives us

THEOREM 3. *Let Σ be the set of sequences f: $\mathbf{N} \to \mathbf{N}$ which satisfy the hypotheses of theorem 2. Then Σ forms a semi-group under composition, whose identity is the sequence e, where $e(n) = n$. Every sequence in Σ has an infinite number of prime divisors.*

By theorem 1 and these remarks, if $f \in \mathbf{Z}[x]$ and $\theta > 0$, $\alpha \geq 1$, then

$$\{g(n)\}_{n=1}^{\infty} = \{f([\theta n^{\alpha}])\}_{n=1}^{\infty}$$

has an infinite number of prime divisors. The same is true of the sequence $\{h(n)\}_{n=1}^{\infty}$, where

$$h(n) = \left[\theta |f(n)|^{\alpha}\right].$$

Along more bizarre lines, the sequence $\{t(n)\}_{n=1}^{\infty}$ defined by

$$t(n) = \left[\theta_1 \left[\theta_2 n^{\alpha_2}\right]^{\alpha_1}\right]$$

lies in Σ if $\theta_1, \theta_2 > 0$ and $\alpha_1, \alpha_2 \geq 1$, as does the sequence

$$\left\{\left[\pi \left[\pi n^2\right]^e\right]\right\}_{n=1}^{\infty}, \quad \text{or} \quad \{s(n)\}_{n=1}^{\infty},$$

where

$$s(n) = \left[\pi\left[\pi^2\left[\pi^3 \cdots \left[\pi^k n^e\right]\right]\right]\right].$$

Exercise (for the skeptical reader): Forget everything you have read in this paper and show that the sequence $\{s(n)\}_{n=1}^{\infty}$ has an infinite number of prime divisors.

We mention one more obvious property of Σ: it is closed under translation. If $f \in \Sigma$ and $k \in \mathbf{N}$, then the sequence f_k defined by

$$f_k(n) = f(n) + k$$

lies in Σ also. More generally, we have

THEOREM 4. *If $a(n) \neq 0$ and $b(n)$ are bounded sequences of integers, and if f satisfies the conditions of theorem 2, then so does g, where*

$$g(n) = a(n)f(n) + b(n).$$

The proof is an easy exercise.

COROLLARY. *If $f \in \Sigma$ and $k \in \mathbf{Z}$, then the congruence*

$$f(n) \equiv k \pmod{p}$$

has a solution n for infinitely many primes p.

3. Extensions to imaginary quadratic fields. But can we extend the original proof? Considering the step at equation (2) we find we can.

THEOREM 5. *Let f: $\mathbf{N} \to \mathbf{Z}$ be a function satisfying*

(a′) *$|f(n)| \leq an^m$ for sufficiently large n and some a, $m > 0$;*

(b′) *$f(n) = k$ has at most ck^ϵ solutions for every integer $k \geq 1$, where c is a constant and $0 < \epsilon < 1/m$.*

Then infinitely many primes divide the terms of $\{f(n)\}_{n=1}^{\infty}$.

Proof. Form the series S as in (1). Then our new hypothesis (b′) implies that

$$S \leq c \sum_{a_1, \ldots, a_k = 0}^{\infty} \frac{\left(p_1^{a_1} \cdots p_k^{a_k}\right)^\epsilon}{p_1^{a_1/m} \cdots p_k^{a_k/m}}$$

$$= c \prod_{r=1}^{k} \left(1 - p_r^{\epsilon - 1/m}\right)^{-1} < \infty,$$

if the only primes dividing $f(n)$ for $n \geq 1$ are $p_1, \ldots, p_k$. However, S is a divergent series by (a′), as before. Thus $\{f(n)\}_{n=1}^{\infty}$ must have an infinite number of prime divisors.

Theorem 5 has the following consequence.

THEOREM 6. *If f and g satisfy the hypotheses of theorem 2 and a, b and c are integers for which $a \neq 0$ and $b^2 - 4ac < 0$, then the sequence $\{h(n)\}_{n=1}^{\infty}$ given by*

$$h(n) = af^2(n) + bf(n)g(n) + cg^2(n)$$

has an infinite set of prime divisors.

This fact depends on the arithmetic in the imaginary quadratic field $\mathbf{Q}(\sqrt{-d})$, where $-d = b^2 - 4ac$. We shall derive it from the following more general result, whose statement and proof use several notions from elementary algebraic number theory, namely, algebraic integers, ideals, and the norm function

$$\text{Norm}\left(x + y\sqrt{-d}\right) = x^2 + dy^2$$

from the field $\mathbf{Q}(\sqrt{-d})$ to $\mathbf{Q}$. Recall that an *algebraic integer* is an algebraic number which satisfies a monic polynomial equation with integral coefficients, and that the set of algebraic integers in the field $\mathbf{Q}(\sqrt{-d})$ forms a ring R_d. In the following proof we make essential use of the fact that R_d has a finite number of units, for any $d > 0$. (See [7].)

THEOREM 7. *Let f: $\mathbf{N} \to R_d$ be any function satisfying*:

(a_d) *Norm $f(n) \leq an^m$ for large n and a, $m > 0$;*

(b_d) $f(n) = \lambda$ *has a bounded number of solutions, for all* $\lambda \in R_d$.

Then the sequence $\{f(n)\}_{n=1}^{\infty}$ *has an infinite number of prime ideal divisors in the ring* R_d.

Proof. We define $g(n) = \text{Norm} f(n)$ and assume that the sequence $\{f(n)\}_{n=1}^{\infty}$ has only a finite number of prime ideal divisors $\mathfrak{p}_i$, $1 \le i \le k$, in R_d. It then follows from the theory of ideals in quadratic fields [3] that for every n, the principal ideal $(f(n))$ is a product of powers of the ideals $\mathfrak{p}_i$:

$$(f(n)) = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_k^{a_k}, \tag{3}$$

and so

$$g(n) = \text{Norm} f(n) = p_1^{f_1 a_1} \cdots p_k^{f_k a_k}, \quad \text{for all} n \ge 1, \tag{4}$$

for some integers a_i and integral primes p_i (not necessarily distinct) defined by Norm $\mathfrak{p}_i = p_i^{f_i}$, $f_i = 1$ or 2.

We obtain a contradiction by showing that g satisfies the hypotheses of theorem 5. Condition (a′) is equivalent to (a_d). To prove (b′) we show that for every $\epsilon > 0$ the number of solutions of

$$g(n) = A \tag{5}$$

is $\le cA^{\epsilon}$, for some constant c. Factoring A into a product of prime powers and using (4) shows that equation (5) is equivalent to

$$p_1^{f_1 a_1} \cdots p_k^{f_k a_k} = A = q_1^{b_1} \cdots q_r^{b_r}, \tag{6}$$

or

$$\sum_{\substack{i \\ p_i = q_j}} f_i a_i = b_j, \quad \text{for} \quad 1 \le j \le r, \tag{7}$$

since several of the primes p_i can equal q_j. Without loss of generality $r \le k$; otherwise (5) has no solutions. By elementary results in the theory of quadratic fields, either only one prime p_i equals q_j, or two primes p_i and p_i' equal q_j, in which case $f_i = f_i' = 1$. Hence (7) becomes

$$a_i = b_j, \quad 2a_i = b_j, \quad \text{or} \quad a_i + a_i' = b_j.$$

In any case (7) has at most $b_j + 1 \le 2b_j$ positive integral solutions in the variables a_i, for each j. It follows that (6) has at most

$$2^r b_1 \cdots b_r \le 2^r \left(\frac{b_1 + \cdots + b_r}{r} \right)^r$$

solutions, by the arithmetic-geometric mean inequality. Finally,

$$b_1 + \cdots + b_r \le b_1 \log_2 q_1 + \cdots + b_r \log_2 q_r = \log_2 A,$$

and so

$$2^r b_1 \cdots b_r \le 2^r \left(\frac{\log_2 A}{r} \right)^r \le 2^k (\log_2 A)^k \le cA^{\epsilon},$$

for any fixed $\epsilon > 0$ and a positive constant c depending on ϵ. This shows that there are at most cA^{ϵ} choices for the a_i in (3). Since the ring R_d has a finite number of units (at most 6), $f(n)$ can take on at most $6cA^{\epsilon}$ values in R_d, for any given A. By assumption, therefore, there are at most $c_1 A^{\epsilon}$ values of n for which (5) holds. This proves (b′), as long as ϵ is chosen less than $1/m$. Theorem 5 now implies that

$\{g(n)\}_{n=1}^{\infty}$ has an infinite number of prime divisors, contradicting (4). This proves the theorem.

To deduce theorem 6 from theorem 7, we write $a \cdot h(n)$ as the norm of a suitable element in $\mathbf{Q}(\sqrt{-d})$, where $-d = b^2 - 4ac$:

$$\begin{aligned} ah(n) &= (af(n) - ag(n)\theta)(af(n) - ag(n)\theta') \\ &= \text{Norm}(af(n) - ag(n)\theta), \end{aligned}$$

where θ and θ' are the roots of $ax^2 + bx + c = 0$. Since $a\theta$ satisfies a monic equation, $a\theta \in R_d$ and theorem 7 applies to the function $af(n) - ag(n)\theta$. The hypotheses of theorem 7 follow easily from (a) and (b) of theorem 2 since 1 and θ are linearly independent over $\mathbf{Q}$.

As an example take $f(n) = [\pi n^2]$, $g(n) = \left[\frac{\sqrt{2}}{en}\right]$, and $a = 1$, $b = -1$, $c = 1$. This gives the sequence $\{c_n\}_{n=1}^{\infty}$ considered at the beginning of the paper.

4. Conclusion and open questions. Our proof does not apply to exponential sequences like

$$\{2^n + 1\}_{n=1}^{\infty} \quad \text{or} \quad \{[10^n\pi]\}_{n=1}^{\infty}, \tag{8}$$

since the corresponding series in (1) is convergent for any m.

The first sequence in (8) can be seen to have an infinite set of prime divisors using the theory of Lucas sequences [**1**], or more advanced results from algebraic number theory (see [**4**], [**6**]). It can also be deduced from a striking theorem of H. Kobayashi [**5**]. This theorem states that if the sequence $\{f(n)\}_{n=1}^{\infty}$ has a finite number of prime divisors, then the translated sequence $\{f(n) + a\}_{n=1}^{\infty}$ has an infinite set of prime divisors, for any nonzero integer a! Taking $f(n) = 2^n$ and $a = 1$ gives the first sequence in (8). Can Kobayashi's theorem be proved in an elementary way? I do not know.

It is not known whether the second sequence in (8) has an infinite set of prime divisors; this sequence consists, of course, of integers which are finite strings of digits from the decimal expansion of π.

In this paper we have only proved positive results, and so we ask a final question (from Paul Bateman) in the opposite direction: is there an elementary proof of the fact that infinitely many primes do *not* divide $\{f(n)\}_{n=1}^{\infty}$, for any irreducible polynomial $f \in \mathbf{Z}[x]$ of degree > 1?

REFERENCES

1. J. Brillhart, D. H. Lehmer, and J. L. Selfridge, New primality criteria and factorizations of $2^m \pm 1$, *Math. Comp.*, 29 (1975) 620–647.
2. I. Gerst and J. Brillhart, On the prime divisors of polynomials, *Amer. Math. Monthly*, 78 (1971) 250–266.
3. H. Hasse, Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper II, Physica Verlag, Würzburg, 1965, p. 138.
4. H. Hasse, Über die Dichte der Primzahlen p, für die eine vorgegebene ganzrationale Zahl a $\neq$ 0 von gerader bzw. ungerader Ordnung mod p ist, *Math. Annalen*, 168 (1966) 19–23.
5. H. Kobayashi, On existence of infinitely many prime divisors in a given set, *Tokyo J. Math.*, 4, No. 2 (1981) 379–380.
6. J. C. Lagarias, The set of primes dividing the Lucas numbers has density 2/3, *Pac. J. Math.*, 118 (1985) 449–461.
7. W. J. LeVeque, Topics in Number Theory, vol. II, Addison-Wesley Publishing Co., 1961.
8. I. Schur, Über die Existenz unendlich vieler Primzahlen in einiger speziellen arithmetischen Progressionen, *Sitzungsberichte Berlin. Math. Gesell.*, II (1912) 40–50.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

Finding the Finite Groups of Symmetries of the Sphere

MARJORIE SENECHAL
Department of Mathematics, Smith College, Northampton, MA 01063

1. Introduction. In his famous *Lectures on the Icosahedron*, Felix Klein [**6**] presented a simple and beautiful derivation of the finite subgroups of the group of rotations of the sphere. His proof, which has been repeated often, is a blend of algebra and geometry ideally suited for a first course in the theory of groups. Indeed, H. S. M. Coxeter received a spontaneous ovation from a largely nonmathematical audience when he presented this proof during his lecture [**3**] at the Shaping Space Conference at Smith College in 1984! The purpose of this note is to present a very simple way of extending the list of rotation groups to include all finite subgroups of symmetries of the sphere, up to conjugation in its full group. In contrast to the standard argument (see section 3), it is a direct continuation of Klein's method, it is intuitively clear, and the list can easily be completed by undergraduate students.

2. The sphere and its symmetries. The sphere in question is the ordinary sphere in three-dimensional Euclidean space $\mathbb{E}^3$, also known as the 2-sphere because it is a two-dimensional manifold. The set of all symmetries of the sphere constitutes its symmetry group, the orthogonal group $O(3)$; the subgroup of orientation-preserving symmetries is the special orthogonal group $SO(3)$. The finite subgroups of $O(3)$ are of interest because they are the symmetry groups of polyhedra and all other "finite figures."

First let us review the finite subgroups of $O(2)$, the symmetry group of the circle. There are two families of groups: cyclic groups **n** generated by n-fold rotation about the circle's center, and dihedral groups **nm** generated by reflection in two lines forming an angle of $180°/n$. This result appears to be very old; Hermann Weyl [**8**] notes that Leonardo da Vinci,

> engaged in systematically determining the possible symmetries of a central building and how to attach chapels and niches without destroying the symmetries of the nucleus.

(An examination of the Codex Ashburnham suggests, however, that Weyl may have read too much into Leonardo's pretty drawings.)

The finite subgroups of $O(3)$ may be said to have been implicitly studied by the ancient Greeks, since they include the symmetry groups of the regular polyhedra, but the first systematic attempts to enumerate all of them seem to have been made by nineteenth-century crystallographers in their efforts to classify crystal forms. They were, however, interested only in those groups that describe ideal crystal forms, polyhedra whose rotational symmetries have orders 1, 2, 3, 4, or 6. These crystallographers used more or less ad hoc symmetry arguments, and sometimes

overlooked the composite isometry, rotary reflection. Klein brought elegance, rigor, and abstraction to the subject, and made it possible to produce a complete list of the groups.

The symmetries of the sphere are the isometries of $\mathbb{E}^3$ that map the sphere onto itself. Isometries are distance-preserving maps; a fixed-point argument shows that in $\mathbb{E}^3$, there are six types: *rotation* about an axis, *reflection* in a plane, *translation* along a vector, *rotary reflection* (rotation followed by reflection in a plane orthogonal to the rotation axis, or vice versa), *screw rotation* (rotation followed by translation in the direction of the rotation axis, or vice versa), and *glide reflection* (reflection followed by translation in the plane of reflection, or vice versa). Since the symmetries of the sphere fix its center, neither simple nor composite translation is among them. Thus the symmetries of the sphere are rotation, reflection, and rotary reflection. Note that when the angle of rotation is 180°, rotary reflection is equivalent to central inversion, and when it is 0°, rotary reflection reduces to reflection.

3. Klein's method. Since rotations are the only orientation-preserving symmetries of the sphere, a finite subgroup H of $SO(3)$ consists entirely of rotations. The rotation axes intersect in a point that is fixed by all the elements of H.

We will briefly review Klein's enumeration; for more details consult [**3**], [**4**], or [**8**]. Let H be a finite subgroup of $SO(3)$ of order N. To each element of H (other than the identity) there corresponds an axis that intersects the sphere in two points. These points are called the *poles* of the rotation. The stabilizer subgroup of any pole is cyclic, being generated by rotation about a single axis. Klein counts, in two different ways, the pairs (h, p), where $h \in H$, p is a point on the sphere, and $hp = p$ (i.e., p is a pole of h).

First, since there are two poles for each of the $N - 1$ nontrivial rotations, the number of pairs is $2(N - 1)$. The second way of counting is a little more involved. The poles fall into k orbits under the action of H; the ith orbit has n_i poles, whose stabilizer subgroups have order s_i. (For each orbit, the stabilizers of the poles are conjugate, and hence all have the same order.) Thus for the ith orbit there are $n_i(s_i - 1)$ pairs (h, p). Now sum over all orbits and equate the two calculations. Using the well-known and easily established fact that, since H is finite, $s_i n_i = N$, we obtain the equation

$$\sum 1/s_i = k - 2 + 2/N.$$

Because each s_i is at least 2, k is less than 4; since N is at least 2, k is greater than 1. Thus H has either 2 or 3 orbits of poles. The few solutions are easily computed; they are shown in Table 1. In that table, the first three columns list the numbers of poles in each of the orbits of H, while the symbol in the fourth column lists the symbols for the cyclic stabilizer subgroups of the rotations corresponding to these orbits. Since H is characterized by the orders of these rotations, the symbol can be used to denote the group itself. In the fifth column we give the order of H, and in the sixth the name of a polyhedron whose rotational symmetry group is H.

For each group in Table 1, the data force the relative positions of the poles on the sphere. Consider, for example, the third line of the table. There are three orbits of poles. The first two consist of four poles each, while the third orbit has six poles. Column four shows that the orders of the corresponding rotation axes are 3, 3, and

n_1	n_2	n_3	H	$\lvert H\rvert$	polyhedron
1	1	—	**nn**	n	n-gonal pyramid
2	n	n	**n22**	$2n$	n-gonal prism or antiprism
4	4	6	**332**	12	regular tetrahedron
6	8	12	**432**	24	cube
12	20	30	**532**	60	regular icosahedron

TABLE 1. The finite subgroups of $SO(3)$. Each subgroup H is characterized by the orbits of the poles of its rotation axes. For each subgroup there are two or three orbits; the sizes of the orbits are given in the first three columns. In the fourth column H is denoted by a symbol which juxtaposes the symbols for the stabilizer subgroups of the poles of each orbit. In the fifth column we give the size of H, and in the sixth an example of a polyhedron whose rotation group is H.

2, respectively. Now, a rotation of order 3 fixing a pole of the first orbit must map the second orbit onto itself; this is possible only if three of its four poles lie equispaced on a circle orthogonal to the rotation axis, and the fourth pole is the other pole of that rotation. It follows that the poles of the first and second orbits must be located at the vertices and face centers, respectively, of a regular spherical tetrahedron, and the poles of the third orbit must be located at the midpoints of the tetrahedron's edges (FIGURE 1). Arguing in this way, we find that the finite rotation groups of the sphere are the infinite families of groups of rotations of the n-gonal pyramid and the n-gonal prism (and antiprism), and the groups of rotations of the regular tetrahedron, the cube, and the regular icosahedron.

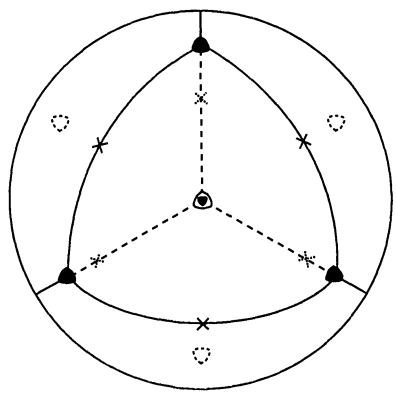


FIG. 1. The poles of a rotation group which belong to 3 orbits of sizes 4 (white triangles), 4 (dark triangles), and 6 ($\times$'s) with corresponding stabilizer subgroups of orders 3, 3, and 2 must be located at the face centers, vertices, and midedges of a regular spherical tetrahedron. Dotted lines indicate that the figure lies on the hemisphere below the plane of the paper.

This completes Klein's enumeration of the finite subgroups of $SO(3)$.

4. The finite subgroups of $O(3)$. We now turn to the main problem of this paper, finding all finite subgroups G of $O(3)$.

If G does not consist of rotations alone, then its rotation subgroup $G \cap SO(3)$, which we will call H, must be of index 2 in G, because the composition of two

orientation-reversing symmetries (of the sphere) is a rotation. Thus $G = H \cup qH$, where the coset representative q is not a rotation.

The argument presented by Pólya and Meyer in [7] seems to have become standard (see, e.g., [**1**, **2**, **8**] and, in more abstract form, [**5**]). The group action argument is abandoned. Instead, two cases are distinguished, according as we may or may not choose q to be inversion through the center of the sphere. In the first case, G is the direct product of H and a cyclic group of order 2. In the second case, q can be taken to be a rotary inversion (see section 5) whose rotational component is *not* an element of H, but of a rotation group which contains H as a subgroup of index 2.

The method works: every finite subgroup of $O(3)$ is either a rotation group or a member of one of these two classes. But, as Johann Kepler remarked apropos Plato's association of the regular solids with the elements, "it has no force of necessity; indeed it admits of other interpretations." There is a very real sense in which distinguishing these two classes is unsatisfactory. Rotary inversion is rather complex geometrically and it is difficult to visualize the construction of the groups when they are derived in this way. Moreover, one would like the symmetry groups of the regular polyhedra to be classed together. But the full groups of symmetries of the cube and the icosahedron contain central inversion, so they belong to the first class, while the full group of symmetries of the tetrahedron does not, so it belongs to the second class. (When central inversion is added to the group of rotations of the tetrahedron, we obtain the symmetry group of the "pyritohedron," a nonregular pentagonal dodecahedron which is an ideal form of the pyrite crystal.)

There is, however, another interpretation. We can derive the finite subgroups of $O(3)$ in a way that complements Klein's derivation of their rotation subgroups. Using this approach, which can be described as a group-theoretic justification of the intuitive method of derivation that crystallographers and other scientists have been using for a long time, students can easily visualize the construction of the groups G and deduce the complete list.

The main idea is the simple observation that, since H is a subgroup of index 2 in G, the isometry q which we add to it to obtain G must permute the poles of H. Either q preserves the ith H-orbit, or it merges it with another orbit of the same size. The latter option is of course available only if a second orbit of the same size exists.

If q preserves an orbit of H, then the order of the stabilizer subgroup of the orbit is doubled in G, since the size of an orbit times the order of its stabilizer subgroup equals the order of the group, and $|G| = 2|H|$. Since the stabilizer subgroup is now dihedral, q can be taken to be a reflection in a plane containing a pole of the ith orbit. When $k = 3$, there is always at least one orbit that must be preserved by q, because it is the only one of its size. Thus q can always be chosen to be a reflection in a suitable plane containing a pole of such an orbit. Note that this choice of q completely determines the extension of the cubic and icosahedral rotation groups: the three orbits of poles of these groups are of different sizes, so q must fix each of them (see FIGURE 2). For the groups **n22** and **332**, there are two choices for q, according as q fixes or permutes the two orbits of the same size (see FIGURE 3). In the first case, we obtain the symmetry groups of the prisms and of the regular tetrahedron; in the second, the symmetry groups of the antiprisms and of the pyritohedron.

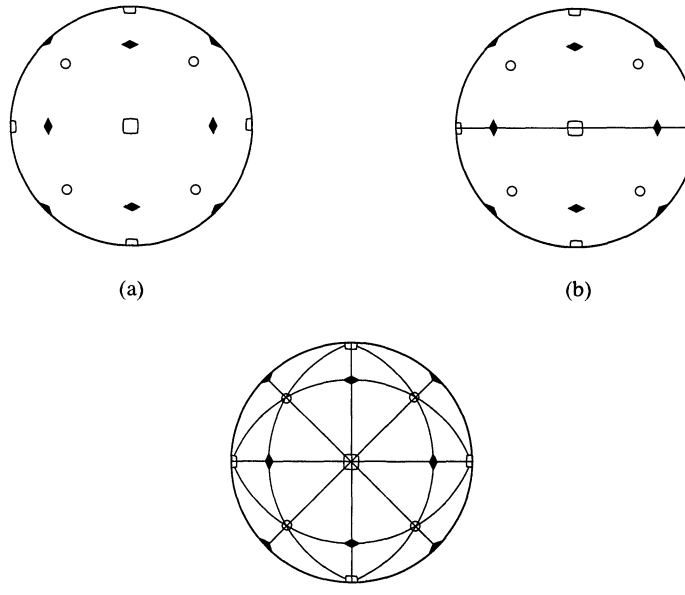


FIG. 2. The group of rotations of the cube (a) can be extended to (c) the full group of cubic symmetries by adding reflection in a plane (b) which maps each of the three orbits of poles onto themselves.

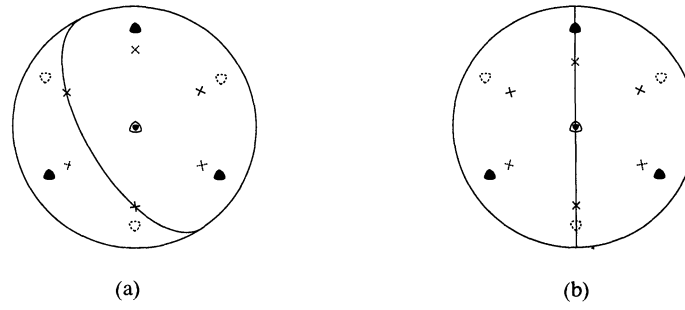


FIG. 3. Here H is the tetrahedral rotation group **332**. (a) q can be chosen to be reflection in a plane which maps the orbits of poles of order 3 onto one another. (b) q can also be chosen so that all orbits are mapped to themselves.

When $k = 2$, H is cyclic. Again, q must fix or interchange the two poles. When q fixes them, the group G is dihedral. When q interchanges them, it is necessary to distinguish the cases when q is a reflection or a rotary reflection q.

The finite subgroups of $O(3)$ (which are not in Table 1) are listed in Table 2.

5. A word about notation. There are (at least) two sources of confusion that one encounters in reading in the literature: the distinction between rotary reflection and rotary inversion, and the seemingly endless proliferation of notation for the groups we have been discussing.

H	G	polyhedron
nn	**nm nm**	n-gonal pyramid
	n / m	—
	$\widetilde{\mathbf{2n}}$	rhomic disphenoid, when $n = 2$
n22	**nm 2m 2m**	n-gonal prism
	nm 2 2	n-gonal antiprism
332	**3m 3m 2m**	regular tetrahedron
	3 3 2m	pyritohedron
432	**4m 3m 2m**	cube
532	**5m 3m 2m**	regular icosahedron

TABLE 2. The finite subgroups of $O(3)$. We extend the list of groups H in Table 1 to include reflections and rotary reflections. H is listed in column 1. The symbol in the second column indicates the G-stabilizer subgroups of the poles of H; the symbol completely characterizes G. (The notation is explained in the text.) G is the symmetry group of the polyhedron in the third column.

Recall that n-fold rotary reflection, usually denoted by $\tilde{n}$, is a composite motion, rotation through $360°/n$ followed (or preceded) by reflection in a plane orthogonal to the rotation axis. This motion generates a cyclic group $\tilde{\mathbf{n}}$ of order n or $2n$, according as n is even or odd. When n is even, the group has orbits consisting of n points arranged alternately above and below the plane of reflection. However, when n is odd, it has orbits of $2n$ points arranged in mirror-image pairs. The latter group can also be generated by a rotation of order n and reflection in a plane orthogonal to the rotation axis. This is the type of group that crystallographers denote by **n/m**, Thus we will use the symbol $\tilde{n}$ only for even values of n.

Rotary inversion, on the other hand, is the composition of n-fold rotation about an axis with central inversion through a point on that axis; it is usually denoted by $\bar{n}$. Rotary inversion is even more confusing than rotary reflection: its properties vary according as $n = 4k$, $n = 4k + 2$, or $n = 4k \pm 1$. Thus $\bar{n} = \tilde{n}$ when n is divisible by 4, $\bar{n} = \widetilde{2n}$ when n is odd, and $\bar{n} = (n/2)/m$ when $n/2$ is odd. Despite this, rotary inversion seems to be more commonly used.

There is no commonly accepted notation for the finite subgroups of $O(3)$, and there are no commonly accepted names for most of them. The notation used in this paper was inspired by crystallographic usage, which, when $k = 3$, denotes the rotation groups by the orders of rotation corresponding to the three orbits of poles. However, I have departed from the crystallographic symbols in order to keep the notation consistent with the derivation of the groups (see Table 3). As explained above, in Table 1 my symbol for each group is simply the juxtaposition of the symbols for the cyclic stabilizer subgroups of poles of the corresponding orbits. This symbolism is continued in Table 2: the symbol for the rotation subgroup is simply augmented, when appropriate, to denote the G-stabilizer subgroup of the pole. For example, the two extensions of the tetrahedral rotation group **332** are denoted **3m3m2m** and **332m**; the first indicates that all three stabilizers are now dihedral, while the second indicates that only the third stabilizer is. However, when $k = 2$ and q interchanges the two poles, both stabilizers remain cyclic. In this case q must be a rotary reflection and we have to distinguish the cases when the angle of rotation is 0 or greater than 0. In the first case G is a group of type **n/m**, and in the second of type $\widetilde{2n}$.

H	IUCr	C & M	Coxeter	G & B	Remark
nn	n	$[n]^+$	C_n	C_3^n	
n22	$n22$	$[2, n]^+$	D_n	H_3^n	$n \geq 2$
332	23	$[3,3]^+$	A_4	T	
432	432	$[3,4]^+$	S_4	W	
532	—	$[3,5]^+$	A_5	I	
nm nm	nmm	$[n]$	D_nC_n	$H_3^n]C_3^n$	
n / m	n/m	$[2, n^+]$	$C_n x\{I\}$	C_3^n*	n even
	$\overline{2n}$		$C_{2n}C_n$	$C_3^{2n}]C_3^n$	n odd
ñ	$(\overline{n/2})$	$[2^+, n^+]$	$C_n x\{I\}$	C_3^n*	n even, $n/2$ odd
	$\bar{n}$		$C_{2n}C_n$	$C_3^{2n}]C_3^n$	n even, $n/2$ even
nm 2m 2m	$n/m\ mm$	$[2, n]$	$D_n x\{I\}$	H_3^n*	n even
	$\overline{2n}m2$		$D_{2n}D_n$	$H_3^{2n}]H_3^n$	n odd, $n > 1$
nm 2 2	$\overline{2n}2m$	$[2^+, 2n]$	$D_{2n}D_n$	$H_3^{2n}]H_3^n$	n even
	$\bar{n}m$		$D_n x\{I\}$	H_3^n*	n odd, $n > 1$
3m 3m 2m	$\bar{4}3m$	$[3,3]$	S_4A_4	$W]T$	
3 3 2m	$m3$	$[3^+, 4]$	$A_4 x\{I\}$	$T*$	
4m 3m 2m	$m3m\ 2/m$	$[3,4]$	$S_4 x\{I\}$	$W*$	
5m 3m 2m	—	$[3,5]$	$A_5 x\{I\}$	$I*$	

TABLE 3. Comparison of notations for the finite subgroups of $O(3)$. In the first column we list the notation used in this paper. In the second, we give the symbol used by the International Union of Crystallography (IUCr); these symbols are defined only for the "crystallographic" groups, whose rotational symmetries have orders 1, 2, 3, 4, or 6. The third column (C & M) lists the symbols used by Coxeter and Moser [4]. The symbols in columns four and five are those used in [2] and [5] by Coxeter and by Grove and Benson (G & B), respectively.

Some of the many notations that can be found in the literature are shown in Table 3. The rationales, advantages, and disadvantages of these different notations could serve as the basis for an interesting and instructive class discussion!

REFERENCES

1. R. P. Burns, Geometry: A Path to Groups, Cambridge University Press, Cambridge, 1987.
2. H. S. M. Coxeter, Introduction to Geometry, Wiley, New York, second ed., 1969.
3. _____, Regular and semi-regular polyhedra, in Shaping Space: A Polyhdedral Approach, M. Senechal and G. Fleck, eds., Birkhauser-Boston, New York, 1988.
4. H. S. M. Coxeter and W. O. J. Moser, Generators and Relations for Discrete Groups, fourth ed., Springer-Verlag, New York, 1984.
5. L. C. Grove and C. T. Benson, Finite Reflection Groups, second ed., Springer-Verlag, New York, 1985.
6. F. Klein, Lectures on the Icosahedron, 1884; see Chap. 5, sec. 2 of the English translation, Dover, N.Y., 1957.
7. G. Pólya and B. Meyer, Sur les symétries des fonctions sphériques de Laplace, *Comptes Rendus Acad. Sci. Paris*, 228 (1949) 28—30.
8. H. Weyl, Symmetry, Princeton University Press, Princeton, 1952.

A Note on Defining Conditional Probability

ARTHUR RAMER
School of Electrical Engineering and Computer Science, University of Oklahoma, Norman, OK 73019

The notion of conditional probability is usually introduced on the basis of frequencies of events belonging to a specified subset (of the set all admissible events) events.

Given $P(x_i) = p_i$, $i = 1, \ldots, n$, not identically 0 on the set of events $\{x_i\}$, $i = 1, \ldots, m$, $m < n$, let us suppose that we wish to define an associated conditional distribution $Q(x_i) = q_i$, $i = 1, \ldots, m$, on that subset. It is then defined as

$$q_i = \frac{p_i}{\sum_{k=1}^{n} p_k},$$

Information theory offers an alternative method of defining a conditional distribution, through the notion of information distance which is interpreted as the difference of the information content between the distributions. This approach suggests that the conditional distribution should be defined as the one for which the distance from the original distribution is minimal.

Such distance, known as I-divergence [**2**, **3**] is defined for discrete distributions as

$$I(Q, P) = q_i \log \frac{q_i}{p_i},$$

given two distributions $p = (p_i)$, $q = (q_i)$, both corresponding to the same set of events $\{x_i\}$. This distance is defined only if $q_i = 0$ whenever $p_i = 0$; however, it is sufficient for our discussion.

To apply the distance formula we must first expand the sought-for distribution $q_1, \ldots, q_m$ to the same domain as the original one $p_1, \ldots, p_n$. An obvious natural extension is $q_1, \ldots, q_m, 0, \ldots, 0$. Now, under the convention $q_i \log q_i/p_i = 0$ for $q_i = 0$, the problem becomes one of minimizing

$$H(q_1, \ldots, q_m) = \sum_{i=1}^{m} q_i \log \frac{q_i}{p_i}$$

subject to

$$z(q_1, \ldots, q_m) = \sum_{i=1}^{m} q_i = 1$$

and

$$q_i \geq 0, \qquad i = 1, \ldots, m.$$

Application of Lagrange multipliers gives

$$\frac{\delta H}{\delta q_i} + \lambda \frac{\delta z}{\delta q_i} = 0 \quad \text{or} \quad \log \frac{q_i}{p_i} + 1 + \lambda = 0,$$

thus suggesting that

$$q_i = \frac{p_i}{\sum_{k=1}^{m} p_k}$$

is the required answer. To verify it we note that (putting $\bar{p} = \sum_{k=1}^{m} p_k$)

$$\sum_{i=1}^{m} q_i \log \frac{q_i}{p_i} = \sum_{i=1}^{m} q_i \log q_i - \sum_{i=1}^{m} q_i \log \frac{p_i}{\bar{p}} - \log \bar{p}.$$

Because the last term is constant, Shannon's inequality [**1**] shows that the difference of the first two terms is nonnegative and equals 0 only when

$$q_i = \frac{p_i}{q_i}, \quad i = 1, \ldots, m.$$

Occasionally, a dual expression $I(P, Q) = p_i \log p_i/q_i$, or a symmetric sum $I(Q, P) + I(P, Q)$, is used as the measure of distance [**2**]. Again, Shannon's inequality shows immediately that its minimum is reached for the same values of q_i.

Therefore, the defining formula for a conditional probability distribution can be motivated by the observation that it is the closest one to the original distribution and concentrated exclusively on the specified subset of the space of events.

REFERENCES

1. J. Aczel and Z. Daroczy, On Measures of Information and Their Characterization, Academic Press, New York, 1975.
2. I. Csiszar, I-Divergence geometry of probability distributions and minimization problems, *Ann. Prob.*, 3 (1975) 146–156.
3. S. Kullback, Information Theory and Statistics, Wiley, New York, 1959.

Continuous Functions and Connected Graphs

C. E. BURGESS
Department of Mathematics, University of Utah, Salt Lake City, Utah 84112

There seems to be a tendency for students in elementary calculus, and also in some more advanced courses, to picture continuous real functions, with the real line $\mathbb{R}$ as their domain, to be those functions over $\mathbb{R}$ that have connected graphs in the plane $\mathbb{R}^2$. A standard counterexample, of course, is the function that is zero for $x = 0$ and $\sin 1/x$ for $x \neq 0$. I contend that this tendency by students to equate the concepts of "continuity" and "connected graphs" is close to a correct equivalence and that the incorrect impression is very likely based on their lack of a clear understanding of such concepts as "limit points," "closed sets," and "connected sets." It is correct to conclude that continuous real functions over $\mathbb{R}$ are those functions over $\mathbb{R}$ whose graphs, in the plane $\mathbb{R}^2$, are both closed and connected. While this follows from the characterization of a topological line as a locally compact metric space that is separated into two connected sets by each of its points [**4, 5**], it is my main purpose in this note to present an elementary proof. Also, I will

include elementary proofs that for a function $f: \mathbb{R} \to \mathbb{R}$ with a closed graph in $\mathbb{R}^2$, "local boundedness" and the "intermediate value property" furnish two other equivalences to the continuity of f. I have not noticed these equivalences in undergraduate textbooks. The proofs presented here are also valid where the domain of f is a closed interval or a closed ray in $\mathbb{R}$.

THEOREM. *If $f: \mathbb{R} \to \mathbb{R}$ has a closed graph in $\mathbb{R}^2$, then the following four conditions are equivalent*:

(a) *f is continuous.*

(b) *f is locally bounded.*

(c) *f has the intermediate value property.*

(d) *The graph of f in $\mathbb{R}^2$ is connected.*

Our procedure will be to show, with three lemmas, that each of (b) and (c) implies (a) and that (d) implies (c). There are well-known easy proofs that (a) implies (b), (c), and (d), so we do not include them here.

LEMMA 1. *If $f: \mathbb{R} \to \mathbb{R}$ is locally bounded and the graph of f is closed in $\mathbb{R}^2$, then f is continuous.*

Proof. Suppose that f is not continuous at $b \in \mathbb{R}$. Then there is a sequence $\{b_n\}$ converging to b such that the sequence $\{f(b_n)\}$ converges to a number $q \neq f(b)$. Thus the sequence of points $\{(b_n, f(b_n)\}$ converges to the point (b, q) in $\mathbb{R}^2$, contradicting the hypothesis that the graph of f is closed.

LEMMA 2. *If $f: \mathbb{R} \to \mathbb{R}$ has the intermediate value property and the graph of f is closed in $\mathbb{R}^2$, then f is continuous.*

Proof. Suppose that f is not continuous at $b \in \mathbb{R}$. Then for some $\varepsilon > 0$, there is a sequence $\{b_n\}$ converging to b such that, for each n, $|f(b_n) - f(b)| > \varepsilon$. For convenience, we assume that each $f(b_n) > f(b) + \varepsilon$. The intermediate value property implies that, for each n, there is a number c_n between b_n and b such that $f(c_n) = f(b) + \varepsilon$. Thus the sequence of points $\{(c_n, f(c_n))\}$ converges to the point $(b, f(b) + \varepsilon)$, contradicting the hypothesis that the graph of f is closed.

LEMMA 3. *If the function $f: \mathbb{R} \to \mathbb{R}$ has a connected graph in $\mathbb{R}^2$, then f has the intermediate value property.*

Proof. It is easy to see that f would have the intermediate value property if, for $a < b$, the set $\{(x, f(x)) | a \leq x \leq b\}$ is connected. So suppose that this set is the union of two mutually separated sets **H** and **K**; that is, neither of the sets **H** and **K** intersects the closure of the other. There are two cases to consider, and each of them contradicts the hypothesis that the graph of f is connected.

Case 1. The points $(a, f(a))$ and $(b, f(b))$ are in **H** and **K**, respectively. This would mean that the graph of f is the union of the two mutually separated sets $\mathbf{H} \cup \{(x, f(x)) | x \leq a\}$ and $\mathbf{K} \cup \{(x, f(x)) | x \geq b\}$.

Case 2. Both of the points $(a, f(a))$ and $(b, f(b))$ are in **H**. This would mean that the graph of f is the union of the two mutually separated sets **K** and $\mathbf{H} \cup \{(x, f(x)) | x \notin [a, b]\}$.

Remarks. It is an easy exercise to show that the continuity of f, in the above theorem, implies that the graph of f is closed [**3**]. Thus there is a redundancy in the part of the theorem where condition (a) implies (b), (c), and (d). However, without

the requirement that the graph of f be closed, the only other valid implication among conditions (a), (b), (c) and (d) is that (d) implies (c), as shown in Lemma 3. We can use the two functions

$$f_1(x) = \frac{1}{x}\sin\frac{1}{x}, \qquad x \neq 0; \qquad f_1(0) = 0,$$

and

$$f_2(x) = 1, \qquad x > 0; \qquad f_2(x) = 0, \qquad x \leq 0,$$

to see that there is no valid implication between (b) and (c) and none between (b) and (d) unless "closed" is included in the hypothesis. The most interesting question in this respect, raised by the referee, is whether (c) implies (d). We furnish a negative answer to this question with a slight change in a function described by F. B. Jones [2]. In Theorems 2 and 5 of his paper, Jones used a Hamel basis [1] for $\mathbb{R}$ to describe a function g: $\mathbb{R} \to \mathbb{R}$ such that the graph of g intersects every horizontal open interval in $\mathbb{R}^2$. So this function has the intermediate value property. (Jones described some other interesting properties of this function that are not needed for our purposes here.) The graph of g, as described by Jones, is connected. However, we can change g slightly as follows to obtain a function f of the desired type that is not connected. Let $\mathbf{C}$ denote the unit circle with center at the origin. For each x such that $(x, g(x)) \in \mathbf{C}$, let $f(x) = g(x) + 3$; otherwise, let $f(x) = g(x)$. Clearly we still have the property that the graph of f intersects every horizontal open interval, so f satisfies the intermediate value property. To see that the graph of f is not connected, notice that it does not intersect $\mathbf{C}$ yet it intersects both components of the complement of $\mathbf{C}$.

REFERENCES

1. G. Hamel, Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung: $f(x + y) = f(x) + f(y)$, *Math. Ann.*, 60 (1905) 459–462.
2. F. B. Jones, Connected and disconnected plane sets and the functional equation $f(x) + f(y) = f(x + y)$, *Bull. Amer. Math. Soc.*, 48 (1942) 115–120.
3. Bert Mendelson, Introduction to Topology, Allyn and Bacon, Inc., Boston, 1962, p. 64.
4. R. L. Moore, Concerning simple continuous curves, *Trans. Amer. Math. Soc.*, 21 (1920) 333–347.
5. _____, Foundations of Point Set Theory, rev. ed., Amer. Math. Soc. Colloq. Publ., vol. 13, Amer. Math. Soc., Providence, R.I., 1962, p. 95.

The l_2 Matrix Norm and the Spectral Radius

ROBERT A. LESLIE
Department of Mathematics, Agnes Scott College, Decatur, GA 30030

A useful theorem in the theoretical development of iterative techniques to solve systems of linear equations in an undergraduate numerical analysis course is [1, p. 419] and [2, p. 21].

THEOREM. *If A is a real, $n \times n$ matrix then $[\rho(A^tA)]^{1/2} = \|A\|_2$, where $\rho(T)$ is the spectral radius of T and $\|\cdot\|_2$ denotes the l_2 matrix norm.*

In [1], the proof of this theorem is omitted as too advanced for undergraduates, whereas the proof given in [2], and referred to in [1], involves several other results

which most undergraduates will not have seen by the time they take a numerical analysis course.

One idea that almost all these students will be familiar with is that of Lagrange multipliers. Here is an approach to this theorem involving Lagrange multipliers and some elementary linear algebra. In addition, several other results follow quickly as corollaries. All norms used are l_2 norms.

Consider the function F: $R^n \to R$ defined by $F(x) = x^tAx$, where A is a real, $n \times n$ matrix and x is a vector in R^n. We wish to find the critical values of F subject to the constraint that $\|x\| = 1$, i.e., $x^t \cdot x = 1$. Since F is very smooth, then on the unit sphere in R^n there clearly is a solution to this maximum/minimum problem. We define G: $R^n \to R$ by $G(x) = x^t \cdot x - 1$ and apply the method of Lagrange multipliers to this constrained maximum/minimum problem for F subject to the condition that $G \equiv 0$. Since $\nabla G = 2x \neq 0$ for $\|x\| = 1$, the method is applicable, and we set $\nabla F = c \cdot \nabla G$, where c is the Lagrange multiplier.

LEMMA 1. $\nabla F(x) = (A^t + A)x$.

Proof. Let $A = (a_{ij})$. Then

$$\frac{\partial F}{\partial x_i} = x^t \cdot \frac{\partial}{\partial x_i}(Ax) + \frac{\partial x^t}{\partial x_i} \cdot (Ax)$$

$$= x^t \cdot A \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} + (0,\ldots,1,\ldots,0) \cdot (Ax), \text{ where each 1 is in position } i$$

$$= \sum_{j=1}^{n} a_{ji}x_j + \sum_{j=1}^{n} a_{ij}x_j = \sum_{j=1}^{n} (a_{ji} + a_{ij})x_j = \left[(A^t + A)x\right]_i. \quad \square$$

Hence, we have the equation that $(A^t + A)x = 2cx$. Let $\lambda = 2c$ and we have that λ is an eigenvalue of $A^t + A$ and x is the corresponding eigenvector. Furthermore, for any such critical vector x, we have

$$A^tx + Ax = \lambda x$$

$$x^tA^tx + x^tAx = x^t\lambda x = \lambda x^tx = \lambda,$$

since $x^tx = 1$. Therefore,

$$(x^tAx)^t + x^tAx = \lambda.$$

But since x^tAx is a real number, $(x^tAx)^t = x^tAx$, so that

$$x^tAx = \tfrac{1}{2}\lambda.$$

Summarizing these results, we have

THEOREM 1. *If A is a real, $n \times n$ matrix, then*

1. $\max_{\|x\|=1} (x^tAx) = \frac{1}{2}\rho(A^t + A)$
2. $\min_{\|x\|=1} (x^tAx) = \frac{1}{2}$ minimum eigenvalue of $A^t + A$.

COROLLARY 1. $\|A\| = [\rho(A^tA)]^{1/2}$.

Proof.

$$\|A\|^2 = \max_{\|x\|=1} \left(\|Ax\|^2\right) = \max_{\|x\|=1} (Ax)^t(Ax) = \max_{\|x\|=1} (x^tA^tAx)$$

$$= \tfrac{1}{2}\rho\left[A^tA + (A^tA)^t\right] = \tfrac{1}{2}\rho[2 \cdot A^tA] = \rho[A^tA]. \quad \square$$

Other results follow quickly. The first may not be known while the second is a classical result in linear algebra. Implicit in the arguments below is the elementary fact that all eigenvalues of a real, symmetric $n \times n$ matrix are real.

COROLLARY 2. *A is positive definite if and only if $A^t + A$ has all positive eigenvalues.*

Proof. From the theorem, $\min_{\|x\|}(x^tAx) = \frac{1}{2}$ minimum eigenvalue of $A^t + A$. $\square$

COROLLARY 3. *If A is a real, symmetric, positive definite matrix then A has all positive eigenvalues.*

Proof. Here $A^t = A$ so, $\frac{1}{2}$ minimum eigenvalue of $A^t + A$ = minimum eigenvalue of A. $\square$

REFERENCES

1. R. Burden and J. D. Faires, Numerical Analysis, 3rd ed., Prindle, Weber & Schmidt, Boston, 1985.
2. J. M. Ortega, Numerical Analysis, A Second Course, Academic Press, New York and London, 1972.

Does

$$\sum_{k=1}^{\infty} \left\{ k \prod_{j=1}^{d_k} \log^{(j)} k \right\}^{-1}$$

converge?

E3382. *Proposed by Geoffrey R. Robinson, UMIST, Manchester, England.*

Suppose that R is a commutative ring with identity and that A is an n by n matrix with entries in R. If $\det A$ is a zero divisor in R, show that there is an n by n matrix B with entries in R such that B is not the zero matrix O but $AB = BA = O$. More specifically, show that B may be chosen as a polynomial in A, i.e., as a finite linear combination of $I, A, A^2, \ldots$ with coefficients from R.

E3383. *Proposed by H. G. Diamond, University of Illinois at Urbana-Champaign.*

Suppose α is a given number in (0,1]. Show that there exists a unique sequence of positive numbers $a_0 < a_1 < a_2 < \cdots$ such that $a_n/n \to 1$ as $n \to \infty$ and

$$\int_{a_n}^{a_{n+1}} \frac{1 + \cos t}{t} dt = \frac{1}{n + \alpha} \quad (n = 0, 1, 2, \ldots).$$

SOLUTIONS OF ELEMENTARY PROBLEMS

An Implicitly Defined Decreasing Function

E 3287 [1988, 763]. *Proposed by David M. Bloom, Brooklyn College, Brooklyn, NY.*

It is easy to see that for each x in $(1, e)$ there is a unique $y = g(x)$ in (e, ∞) such that $x^y = y^x$, or equivalently $(\log y)/y = (\log x)/x$. For example, $g(2) = 4$. Prove that if $h(x) = x^{g(x)}$, then h is a decreasing function on $(1, e)$.

Solution I by O. P. Lossers, University of Technology, Eindhoven, The Netherlands. Showing that $h(x)$ is decreasing is equivalent to showing that $\operatorname{loglog} h$ is a decreasing function of $\log x$. Writing $u = \log x$ and $v = \log g$, the problem amounts to showing that $v + \log u$ is a decreasing function of u for $u \in (0, 1)$.

Upon taking logs twice, the defining relationship $x^{g(x)} = g(x)^x$ becomes $v + \log u = u + \log v$, which yields $dv/du = (uv - v)/(uv - u)$ under implicit differentiation. We seek to show that

$$\frac{d}{du}(v + \log u) = \frac{uv - 1}{u(v - 1)}$$

is negative for $u \in (0, 1)$. Since $g(x) > e$, we have $v = \log g > 1$, so the denominator is positive and it suffices to show $uv < 1$ for $u \in (0, 1)$.

For this, we use the parametric representation $(u, v) = (t, te^t)/(e^t - 1)$, $0 < t < \infty$, which can be verified by substitution. This yields

$$uv = \left(\frac{t}{e^{t/2} - e^{-t/2}}\right)^2 = \left(\frac{t}{2\sinh(t/2)}\right)^2 < 1.$$

Solution II by Zachary Franco, University of California, Berkeley, CA. Let $y = g(x)$. We parametrize the curve (x, y) by $x = s^{1/(s-1)}$ and $y = s^{s/(s-1)}$ for $s > 1$ and $x = y = e$ for $s = 1$. Note that $s > 1$ implies $y > x$, so that $x^y = y^x$ and $s > 1$ imply $0 < x < e$. Since $y = xs$, we have

$$\frac{dy}{dx} = s + x\frac{ds}{dx}.$$

From $x^{s-1} = s$, we obtain

$$\frac{ds}{dx} = \frac{s(s-1)}{x(1 - s\log x)},$$

so

$$\frac{dy}{dx} = s\frac{s - s\log x}{1 - s\log x}.$$

Now consider $h(x)$; we seek to show $h'(x) < 0$. Since $h(x) > 0$, it suffices to show

$$\frac{h'(x)}{h(x)} = \frac{d\log h}{dx} < 0.$$

We have $\log h = y\log x$, so it suffices to show that

$$\frac{y}{x} + \frac{dy}{dx}\log x = s + s\frac{s - s\log x}{1 - s\log x}\log x = s\frac{1 - s(\log x)^2}{1 - s\log x}$$

is negative. Since $x^s = y > e$, the denominator of the last expression is negative, and it suffices to show $s(\log x)^2 = s(\log s)^2/(s-1)^2 < 1$ for $s > 1$. This is equivalent to $\log s < s^{1/2} - s^{-1/2}$, which follows from the integral $\int_1^s (2t)^{-1}(t^{1/4} - t^{-1/4})^2\,dt > 0$.

Editorial comment. Michael Vowe remarked that the parametrization in the last solution was due to D. Bernoulli. Several readers mentioned the related problem #1 on the 1961 Putnam exam and its solution in Gleason, Greenwood, and Kelly, *The William Lowell Putnam Mathematical Competition* (1980).

Solved also by the proposer and 26 other readers.

Greatest Common Divisors, Revisited

E 3288 [1988, 763]. *Proposed by John J. Cade, Pikeville College, Pikeville, KY.*

It is well known that if a, m, and n are positive integers with $a > 1$, then

$$(a^m - 1, a^n - 1) = a^{(m,n)} - 1,$$

where (m, n) is the greatest common divisor of m and n. Determine $(a^m - 1, a^n + 1)$ and $(a^m + 1, a^n + 1)$.

Solution by Kee-Wai Lau, Hong Kong. Let 2^i and 2^j be the highest powers of 2 dividing m and n, respectively. Then the desired values d and d' are

$$d = (a^m + 1, a^n + 1) = \begin{cases} a^{(m,n)} + 1 & i = j \\ 1, & i \neq j \quad \text{and } a \text{ even}, \\ 2, & i \neq j \quad \text{and } a \text{ odd} \end{cases}$$

$$d' = (a^m - 1, a^n + 1) = \begin{cases} a^{(m,n)} + 1, & i > j \\ 1, & i \leqslant j \quad \text{and } a \text{ even}. \\ 2, & i \leqslant j \quad \text{and } a \text{ odd} \end{cases}$$

Since the techniques for obtaining d and d' are similar, we prove only the result for d (see editorial comment below). Let $b = a^{(m,n)}$, $r = m/(m, n)$, and $s = n/(m, n)$, so that $d = (b^r + 1, b^s + 1)$ and $(r, s) = 1$. Hence there exist positive integers e and f such that $|er - fs| = 1$.

If $i = j$, then r and s are odd. By symmetry we may assume that $er - fs = 1$, e is odd, and f is even. From $b^r \equiv b^s \equiv -1 \bmod d$ we have $(b^r)^e = ud - 1$ and $(b^s)^f = vd + 1$, where u and v are integers. Thus $ud - 1 = (vd + 1)b$, which implies $d|(b + 1)$. Since r and s are odd, $(b + 1)|(b^r + 1)$ and $(b + 1)|(b^s + 1)$, which implies that $d = b + 1$.

If $i \neq j$, then r and s have different parity. By symmetry we may assume that r is even and s is odd. Then $(b^r)^s = yd - 1$ and $(b^s)^r = zd + 1$, where y and z are integers. Hence $yd - 1 = zd + 1$, which implies $d|2$. Since $2|d$ only if a is odd, the result follows.

Editorial comment. Allan Pedersen and O. P. Lossers noted that

$$(a^m - 1, a^n + 1) = (a^m + a^n, a^n + 1) = (a^{|m-n|} + 1, a^n + 1),$$

so that the formula for d' can be obtained from the formula for d.

As David Callan pointed out, the computation of d was problem E 2378 [1972, 906], proposed by D. E. Penney. The solution of E 2378 by Alan Stein [1973, 1059] can easily be extended, as Callan did, to compute d' also. A related problem was mentioned by E. T. H. Wang: If $(a, b) = 1$, then $(a^m - b^m, a^n - b^n) = a^k - b^k$ (proposed by Tom Apostol, *Mathematics Magazine* 53(1980) 49 and 54(1981)86–87). This problem, as well as E 3288, was generalized in the solution of E 3288 by Donald Knuth. Knuth's generalization replaces 1 by powers of b at appropriate places. Let $d' = (a^m - b^m, a^n + b^n)$ and $d = (a^m + b^m, a^n + b^n)$. Knuth proves that $d' = a^{(m,n)} + b^{(m,n)}$ if a higher power of 2 divides m than divides n and $d' = (a + b, 2)$ otherwise, and he proves that $d = a^{(m,n)} + b^{(m,n)}$ if the highest power of 2 dividing m and n is the same and $d = (a + b, 2)$ otherwise.

Solved also by R. Breusch, D. Callan, J. Duemmel, R. B. Eggleton (Brunei), J. Ferrer (Spain), J. Herzog (West Germany), W. Janous (Austria), H.-S. Ki (South Korea), D. E. Knuth, L. Kuipers (Switzerland), J. R. Kuttler, O. P. Lossers (The Netherlands), R. Martin (West Germany), A Nijenhuis, A. Pedersen (Denmark), M. Polakovic (Czechoslovakia), H. Schmidt, Jr., J. H. Steelman, Stanford Ordinal Group, and the proposer.

Two Very Special Mappings

E 3289 [1988, 872]. *Proposed by C. Alsina, Universitat Politecnica de Cataluna, Barcelona.*

Let $I = [0,1]$. Prove that there are exactly two mappings T from $I \times I$ to I that have the following five properties for all x, y, z in I:

(1) $T(T(x, y), z) = T(x, T(y, z))$,
(2) $T(x, y) = T(y, x)$.
(3) $T(x, z) \leqslant T(y, z)$ if $x \leqslant y$,
(4) $T(x, 1) = x$,
(5) $T(zx, zy) = z^k T(x, y)$ for some fixed positive k.

Solution by Robert C. Williams, Alfred University, Alfred, NY. It is easy to confirm that the mappings $T(x, y) = \min(x, y)$ and $T(x, y) = xy$ satisfy the given conditions. We now prove they are the only such mappings.

Using conditions (2), (5), and (4), we have $T(x,0) = T(0, x) = x^k T(0,1) = x^k 0 = 0$. Together with $T(x,1) = x$, this determines T on the boundary of $I \times I$. Now suppose $0 < x \leqslant y < 1$ and apply the same three properties:

$$T(x, y) = T(y, x) = y^k T(1, x/y) = y^k (x/y) = xy^{k-1}.$$

In other words, $T(u, v) = \min(u, v)[\max(u, v)]^{k-1}$ for all (u, v) in the interior of the square. Since this also holds on the boundary, it is valid on all of $I \times I$.

Finally, it suffices to show that the parameter k must be 1 or 2. Choose any y in $[1/2, 1]$ and any x such that $0 < x \leqslant (1/2)\min\{1, y^{k-1}, 2^k y\}$. From (1) we have

$$T\left(T\left(x, \frac{1}{2}\right), y\right) = T\left(x, T\left(\frac{1}{2}, y\right)\right). \qquad (*)$$

Since $x \leqslant 1/2 \leqslant y$, we have $T(x, 1/2) = x(1/2)^{k-1}$ and $T(1/2, y) = y^{k-1}/2$. Inserting these into $(*)$ and using $x \leqslant 2^{k-1} y$ on the left and $x \leqslant y^{k-1}/2$ on the right, the equality becomes

$$x\left(\frac{1}{2}\right)^{k-1} y^{k-1} = x\left(y^{k-1}/2\right)^{k-1}.$$

This reduces to $k - 1 = (k-1)^2$, which implies k is 1 or 2.

Note that condition (3) was not used in the argument, which proves that it is superfluous.

Solved also by the proposer and 78 other readers.

An Inequality for Intertwined Exponentials

E 3291 [1988, 872]. *Proposed by Jerrold R. Griggs, University of South Carolina, Columbia SC.*

Suppose $y > x > 0$. Is it always true that $y^{x^y} > x^{y^x}$?

Solution by Byeong Soo Kim (student), Yonsei University, Seoul, Korea. Yes. It suffices to show that $x^y \ln y - y^x \ln x > 0$. If $x^y \geqslant y^x$, then $y > 1$ and hence

$x^y \ln y - y^x \ln x \geq y^x \ln y - y^x \ln x > 0$. If $x^y < y^x$ and $0 < x < 1$, then

$$x^y \ln y - y^x \ln x > x^y \ln x - y^x \ln x = (x^y - y^x) \ln x > 0.$$

Hence we may assume $x^y < y^x$ and $x \geq 1$. Since $y \ln x < x \ln y$, we have $x^y \ln y - y^x \ln x > (x^y y - x y^x)(\ln x)/x$, and it suffices to show that $x^y y - x y^x \geq 0$. For this it suffices to show $\ln(x^y y) \geq \ln(x y^x)$, so let

$$f(y) = \ln(x^y y) - \ln(x y^x) = y \ln x + \ln y - \ln x - x \ln y.$$

Since f is 0 at x, it suffices to show $f'(y) \geq 0$ for $y > x$. But

$$f'(y) = \ln x + (1 - x)/y \geq \ln x + (1 - x)/x = x^{-1} \int_1^x \ln t \, dt \geq 0.$$

Solved also by J. Anglesio (France), G. Baloglou, R. K. Bose, D. Callan, J. Dalbec, Z. Füredi (Hungary) and the proposer, S. K. Ghosh (India), R. A. Groeneveld, K.-W. Lau (Hong Kong), J. McHugh, T. S. Norfolk, R. Schramm (Israel), R. Stong, N. S. Thornber, and M. Vowe (Switzerland). Partially solved by A. Lenskold and by I. Szalkai (Hungary). One incorrect solution was submitted.

Open Subsets of Closed Sets

E 3294 [1988, 873]. *Proposed by Daniel Goffinet, Saint Etienne, France.*

Characterize the topological spaces in which the following assertion is true: For each open subset G of a closed subset F there exists a set A with G as its interior and F as its closure.

Solution by Jesús Ferrer, Universidad Complutense, Madrid, Spain. We show that a topological space X satisfies the required property if and only if X has a dense subset with empty interior. For necessity, simply take $F = X$ and $G = \varnothing$, and the stated property ensures the existence of such a subset.

For sufficiency, let D be an interior-empty dense subset of X. For each open subset G of a closed subset F, let $A = G \cup (F - \text{int}(F)) \cup (D \cap F - G^*)$, where $\text{int}(Y)$ and Y^* denote the interior and closure of Y. We obtain $A^* = G^* \cup (F - \text{int}(F)) \cup (D \cap F - G^*)^*$. For every $x \in F$ such that $x \in \text{int}(F) - G^*$ and every open neighborhood V of x, it follows that $V \cap (D \cap F - G^*) \supset (V \cap \text{int}(F) - G^*) \cap D \neq \varnothing$. Thus $x \in (D \cap F - G^*)^*$, and $A^* = F$.

Finally, for every $x \in \text{int}(A)$, we have $x \in G \cup (D \cap F - G^*)$. If $x \in D \cap F$, let V be an open neighborhood of x contained in $\text{int}(A)$. It follows that

$$V = V \cap A = (V \cap G) \cup (V \cap (D \cap F - G^*))$$

and $V \cap G \neq \varnothing$, since otherwise $V \subset D$. Thus $x \in G^*$, and $\text{int}(A) \cap (D \cap F - G^*) = \varnothing$, so $\text{int}(A) = G$.

Editorial comment. The stated condition is a natural one, since Euclidean spaces satisfy it. S. F. Barger and Z. Piotrowski point out that there is a literature on the further analysis of this condition, going back to E. Hewitt, *Duke Math. J.* 10(1943) 309–333. K. Kunen, A. Szymański, and F. Tall relate it to measurable cardinals in *Ann. Math. Silesianae* (2) 14(1986) 98–107.

Solved also by S. F. Barger & Z. Piotrowski, S. Beslin & V. Holmes, D. Callan, O. P. Lossers (The Netherlands), C. H. Scott, J. H. Steelman, R. Stong, and the proposer. One incorrect solution was received.

A Binary Summation

E 3303 [1989, 54]. *Proposed by Donald E. Knuth, Stanford University, Stanford, CA.* Suppose m and n are positive integers. Find a closed-form expression for

$$\sum_{k=0}^{2^{mn}-1} \binom{k^m}{n} (-1)^{v(k)},$$

where $v(k)$ is the number of 1's in the binary representation for k.

Solution I by José Heber Nieto, Maracaibo, Venezuela. The value of the sum is

$$(-1)^{nm} \frac{(nm)!}{n!} 2^{nm(nm-1)/2}.$$

Let $S(n, r) = \sum_{k=0}^{2^n-1} k^r (-1)^{v(k)}$. We first prove that $S(n, r) = 0$ for $n > r \geqslant 0$. This is obvious for $r = 0$. Suppose it is true for $0 \leqslant r < t$ and consider $n > t$. Let $\alpha = 2^{n-1}$ for clarity, so $S(n-1, r) = \sum_{k=0}^{\alpha-1} (-1)^{v(k)} k^r$. Then

$$S(n, t) = S(n-1, t) + \sum_{k=\alpha}^{2\alpha-1} k^t (-1)^{v(k)} = S(n-1, t) - \sum_{k=0}^{\alpha-1} (k+\alpha)^t (-1)^{v(k)}$$

$$= S(n-1, t) - S(n-1, t) - \sum_{k=0}^{\alpha-1} p(k)(-1)^{v(k)},$$

where $p(k)$ is a polynomial in k of degree less than t. By the induction hypothesis, this evaluates to 0.

We next find a recursive formula for $S(n, n)$. Breaking this sum into even and odd parts, we have

$$S(n, n) = \sum_{k=0}^{\alpha-1} (2k)^n (-1)^{v(2k)} + \sum_{k=0}^{\alpha-1} (2k+1)^n (-1)^{v(2k+1)}$$

$$= 2^n S(n-1, n) - 2^n S(n-1, n) - \sum_{k=0}^{\alpha-1} \left(n 2^{n-1} k^{n-1} + q(k)\right)(-1)^{v(k)},$$

where $q(k)$ is a polynomial in k of degree less than $n-1$. The result of this computation is $S(n, n) = -n 2^{n-1} S(n-1, n-1)$.

From this we obtain $S(n, n) = (-1)^n n! 2^{n(n-1)/2}$. Since $\binom{k^m}{n} = k^{mn}/n! + r(k)$, where $r(k)$ is a polynomial in k of degree less than nm, we conclude that

$$\sum_{k=0}^{2^{mn}-1} \binom{k^m}{n} (-1)^{v(k)} = \frac{S(nm, nm)}{n!} = (-1)^{nm} \frac{(nm)!}{n!} 2^{nm(nm-1)/2}.$$

Solution II by the proposer. The solution is stated above and is a special case of a formula in L. C. Hsu, *Fibonacci Quarterly*, **11**(1973) 480–484. Hsu proved that if $f(x) = a_N x^N + \cdots + a_1 x + a_0$ is any polynomial of degree at most N, then

$$\sum_{0 \leqslant \varepsilon_1, \ldots, \varepsilon_N \leqslant 1} (-1)^{\varepsilon_1 + \cdots + \varepsilon_N} f(\varepsilon_1 x_1 + \cdots + \varepsilon_N x_N) = (-1)^N N! a_N x_1 \ldots x_N.$$

Setting $N = nm$, $f(x) = \binom{x^m}{n}$, and $x_k = 2^{k-1}$ yields the desired result. The formula

of Hsu given above is easily proved by induction on N. (We apply the formula for $N-1$ to the polynomial $f(x)-f(x+x_N)=-a_N N x_N x^{N-1}+\cdots$.)

Editorial comment. Eugene Levine and Adnah G. Kostenbauder found a generalization of Knuth's formula to base b. Let ω be a primitive bth root of unity and let $v_b(k)$ be the sum of the digits of the base b representation of k. Then

$$\sum_{k=0}^{b^{mn}-1}\binom{k^m}{n}\omega^{v_b(k)}=\frac{(mn)!b^{mn(mn+1)/2}}{n!(\omega-1)^{mn}}.$$

This follows from the corresponding generalization of Hsu's formula:

$$\sum_{0\leqslant\varepsilon_1,\ldots,\varepsilon_N\leqslant b-1}\omega^{\varepsilon_1+\cdots+\varepsilon_N}f(\varepsilon_1x_1+\cdots+\varepsilon_Nx_N)=\left(\frac{b}{\omega-1}\right)^N N!a_Nx_1\cdots x_N.$$

David Callan noted that the case $n=1$ of the original problem was Putnam problem B-5 in 1984.

Solved also by D. Callan, M. Davidson, M. N. Kolountzakis (Greece), A. G. Kostenbauder, V. S. Miller, E. Levine, P. Ševera & O. Šuch, and N. Strauss (Brasil).

Maximizing Symmetric Functions of Side Differences in Triangles

E 3305 [1989, 55]. *Proposed by M. S. Klamkin, University of Alberta, Edmonton.*

If a, b, c are the sides of a triangle with given semiperimeter s, determine the maximum values of

(i) $(b-c)^2+(c-a)^2+(a-b)^2$,
(ii) $|(b-c)(c-a)|+|(c-a)(a-b)|+|(a-b)(b-c)|$, and
(iii) $(b-c)^2(c-a)^2(a-b)^2$.

Solution by David Callan, University of Bridgeport, Bridegport, CT. The maximum values are $2s^2$, s^2, $s^6/108$, respectively. Let a, b, c denote the sides in decreasing order. Note that $s/2\leqslant b\leqslant s$. Form a new (degenerate) triangle with the same perimeter and middle side, the new sides in decreasing order being $s, b, s-b$. In switching to the new triangle, no difference between side lengths decreases; in fact, each difference increases if the original triangle is non-degenerate. Hence each part of the problem is the maximization of a polynomial in b over $[s/2, s]$. The polynomials are:

(i) $(2b-s)^2+b^2+(s-b)^2$,
(ii) $b(2b-s)+b(s-b)+(s-b)(2b-s)$, and
(iii) $(2b-s)^2b^2(s-b)^2$.

These maximizations can be done by elementary calculus, giving maxima as listed above, occurring at $b=s, s, (3+3^{1/2})s/6$, respectively.

Editorial comment. Most solvers treated the problem as a constrained maximization problem in several variables.

Solved also by J. Anglesio (France), E. Arnold. M. Barr (Canada), J. Fukuta (Japan), H. Guggenheimer, W. Janous (Austria), V. Konecný, L. Kuipers, H. Lipman, O. P. Lossers (The Netherlands), A. Pedersen (Denmark), I. A. Sakmar (Turkey), J. H. Steelman, M. Vowe (Switzerland), Western Maryland College Problem Group, and the proposer. Partially solved by V. Schindler (E. Germany) and S. Arslanagić (Yugoslavia). Comments were received from two other readers.

ADVANCED PROBLEMS

6628. *Proposed by Robert A. Melter, Long Island University, Southampton, NY.*

Call a triangle a Heron Triangle if it has integer sides and integer area. Fermat showed that there does not exist a Heron right triangle whose area is a perfect square. However, the triangle with sides 9, 10, 17 has area 36.

Prove that there are infinitely many Heron triangles whose sides have no common factor and whose area is a perfect square.

6629. *Proposed by W. E. Briggs, University of Colorado, Boulder, CO.*

Let σ_i be the ith elementary symmetric function of n positive numbers.

(i) Prove that

$$3(\sigma_{i-1}\sigma_{i+2} + \sigma_i\sigma_{i+1})^2 \geqslant 4(\sigma_{i-1}\sigma_{i+1}^3 + \sigma_{i-1}^2\sigma_{i+2}^2 + \sigma_i^3\sigma_{i+2})$$

for $1 \leqslant i \leqslant n - 1$.

(ii)* Is it true that

$$\sigma_{i-1}\sigma_{i+2}^2 + \sigma_i^2\sigma_{i+3} + \sigma_{i+1}^3 > \sigma_{i+1}(\sigma_{i-1}\sigma_{i+3} + 2\sigma_i\sigma_{i+2})$$

for $1 \leqslant i \leqslant n - 1$?

(iii)* Is it true that

$$\sigma_i^2(\sigma_i^2 - \sigma_{i-1}\sigma_{i+1}) > \sigma_{i-1}^2(\sigma_{i+1}^2 - \sigma_i\sigma_{i+2})$$

for $1 \leqslant i \leqslant n - 1$?

6630. *Proposed by David Pagus, Portland, OR.*

Suppose that $a_1, a_2, a_3, \ldots$ are positive real numbers such that

$$\sum_{n=1}^{\infty} a_n \log(1 + 1/a_n) < \infty.$$

(i) If $\{b_n\}_{n=1}^{\infty}$ is an arbitrary sequence of real numbers, prove that the series

$$\sum_{n=1}^{\infty} a_n|x - b_n|^{-1}$$

converges for almost all real x.

(ii) If k is a fixed positive integer and $\{\vec{b}_n\}_{n=1}^{\infty}$ is an arbitrary sequence of vectors in $\mathbb{R}^k$, prove that the series

$$\sum_{n=1}^{\infty} a_n\|\vec{x} - \vec{b}_n\|^{-k}$$

converges for almost all vectors $\vec{x}$ in $\mathbb{R}^k$. Here $\|\ \|$ is the usual Euclidean norm on vectors in $\mathbb{R}^k$.

SOLUTIONS OF ADVANCED PROBLEMS

Odd Integers with Relatively Large Divisor Sum

6555 [1987, 800]. *Proposed by Moshe Laub, Jerusalem, Israel.*

Let B be the set of natural numbers n such that $\sigma(2n+1) \geq \sigma(2n)$, where $\sigma(k)$ is the sum of the positive integral divisiors of k. Let $\beta(x)$ be the number of elements of B not exceeding x. Prove that there exist constants λ and μ with $0 < \lambda < \mu < 1$ such that

$$\lambda x < \beta(x) < \mu x$$

for all large positive x.

Solution by L. E. Mattics, University of South Alabama, Mobile. We show that $x/3000 < \beta(x) < 25x/28$ for all sufficiently large x. We use the following lemma.

LEMMA. *If h and k are given coprime positive integers and $x \geq 2$, then*

$$\sum_{n \equiv h (\mathrm{mod}\, k),\ n \leq x} \frac{\sigma(n)}{n} = \frac{\pi^2}{6}\frac{x}{k}\prod_{p|k}\left(1 - \frac{1}{p^2}\right) + O(\log x).$$

Proof. If χ is a character modulo k, then

$$\begin{aligned}\sum_{n \leq x} \chi(n)\sigma(n)/n &= \sum_{n \leq x} \chi(n) \sum_{d|n} 1/d \\ &= \sum_{d \leq x} 1/d \sum_{n \leq x,\ n \equiv 0 (\mathrm{mod}\, d)} \chi(n) \\ &= \sum_{d \leq x} \chi(d)/d \sum_{m \leq x/d} \chi(m).\end{aligned}$$

If χ is non-principal, this gives

$$\sum_{n \leq x} \chi(n)\sigma(n)/n = O\left(\sum_{d \leq x} 1/d\right) = O(\log x).$$

If χ is the principal character modulo k and if a dash indicates that a summation is restricted to integers relatively prime to k, we have

$$\begin{aligned}\sum_{n \leq x} \chi(n)\sigma(n)/n &= \sum_{d \leq x}{}' \frac{1}{d} \sum_{m \leq x/d}{}' 1 \\ &= \sum_{d \leq x}{}' \frac{1}{d}\left\{\frac{\phi(k)}{k}\frac{x}{d} + O(1)\right\} \\ &= \frac{\phi(k)}{k} x\left\{\sum_{d=1}^{\infty}{}' \frac{1}{d^2} + O\left(\frac{1}{x}\right)\right\} + O(\log x) \\ &= x\frac{\phi(k)}{k}\frac{\pi^2}{6}\prod_{p|k}\left(1 - \frac{1}{p^2}\right) + O(\log x).\end{aligned}$$

Multiplying these results by $\bar{\chi}(h)$ and summing over all the residue characters χ

modulo k, we obtain

$$\sum_{n\le x,\ n\equiv h(\mathrm{mod}\ k)} \frac{\sigma(n)}{n} = \frac{1}{\phi(k)} \sum_{\chi} \bar{\chi}(h) \sum_{n\le x} \chi(n) \frac{\sigma(n)}{n}$$

$$= x \frac{\pi^2}{6k} \prod_{p|k} \left(1 - \frac{1}{p^2}\right) + O(\log x).$$

Thus the lemma is proved.

To get an upper bound for $\beta(x)$ we take $k = 2$ in the lemma and obtain

$$\sum_{0\le i\le x} \frac{\sigma(2i+1)}{2i+1} = \sum_{n\equiv 1(\mathrm{mod}\ 2),\ n\le 2x+1} \frac{\sigma(n)}{n} = \frac{\pi^2}{8} x + O(\log x).$$

Since $\pi^2 < 10$, this implies that there is an N such that

$$\sum_{1\le i\le x} \frac{\sigma(2i+1)}{2i+1} < \frac{5}{4} x - 2 \quad \text{for } x > N.$$

Since $\sigma(2i+1) < \sigma(2i)$ for $2 \le i \le 6$, we note that if $i \in B$ and $i \ne 1$, then

$$\frac{\sigma(2i+1)}{2i+1} \ge \frac{\sigma(2i)}{2i} \frac{2i}{2i+1} \ge \frac{3}{2} \frac{14}{15} = \frac{7}{5}.$$

Hence

$$\frac{7}{5}\{\beta(x) - 1\} < \sum_{1\le i\le x} \frac{\sigma(2i+1)}{2i+1} < \frac{5}{4} x - 2,$$

so that

$$\beta(x) < 25x/28 \quad \text{for } x > N.$$

We get a lower bound for $\beta(x)$ by considering the intersection of B with the arithmetic progression $n \equiv 7(\mathrm{mod}\ 30)$. Let $\alpha(x)$ be the number of natural numbers n such that $n \le x$, $n \in B$, and $n \equiv 7(\mathrm{mod}\ 30)$, and suppose an asterisk indicates that a summation is restricted to the integers congruent to 7 modulo 30. The lemma gives

$$\sum_{1\le i\le x}^{*} \frac{\sigma(2i)}{2i} = \frac{3}{2} \sum_{1\le i\le x}^{*} \frac{\sigma(i)}{i}$$

$$= \frac{3}{2} \frac{x}{30} \frac{\pi^2}{6} \left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right)\left(1 - \frac{1}{5^2}\right) + O(\log x)$$

$$= \frac{4\pi^2}{25} \frac{x}{30} + O(\log x).$$

Since $\pi^2 < 99/10$, this implies that there is an M such that

$$\sum_{1\le i\le x}^{*} \frac{\sigma(2i)}{2i} < \frac{198}{125} \frac{x}{30} - 4 \quad \text{for } x > M.$$

Now if $i \equiv 7 (\mathrm{mod}\, 30)$, then $2i + 1$ is divisible by 3 and 5 and so, if i is *not* in B,

$$\frac{\sigma(2i)}{2i} > \frac{\sigma(2i+1)}{2i+1} \geq \frac{4}{3}\frac{6}{5} = \frac{8}{5}.$$

Thus if $x > M$, we have

$$\frac{8}{5}\left\{\frac{x}{30} - 2 - \alpha(x)\right\} < \sum_{1 \leq i \leq x}^{*} \frac{\sigma(2i)}{2i} < \frac{198}{125}\frac{x}{30} - 4$$

and hence

$$\frac{8}{5}\left\{\frac{x}{30} - \alpha(x)\right\} < \frac{198}{125}\frac{x}{30},$$

so that $\beta(x) \geq \alpha(x) > x/3000$.

Editorial comment. Adolf Hildebrand remarks that, using the methods of probabilistic number theory, one can show that $\beta(x)/x$ has a limit as $x \to \infty$.

Solved also by P. Erdös and C. A. Spiro jointly.

Expansion by an Iterated Mapping

6566 [1988, 150]. *Proposed by Jonathan E. Spingarn, Georgia Institute of Technology, Atlanta.*

Let $r: R^2 \to R^2$ be a rotation about the origin by an irrational multiple of π. Let $f: R^2 \to R^2$ be defined by

$$f(x, y) = (x, y) \quad \text{if } x \leq 1, \qquad f(x, y) = (2 - x, y) \quad \text{if } x > 1.$$

Does there exist $(x_0, y_0) \in R^2$ such that

$$\|(r \circ f)^k(x_0, y_0)\| > 1$$

for all positive integers k? (Here $(r \circ f)^k$ denotes the kth iterate of $r \circ f$ and $\| \ \|$ denotes the ordinary Euclidean norm.)

Solution by Nicholas Passell, University of Wisconsin, Eau Claire. Yes, there are many such points. In particular, let C be a circle of radius R such that $1 < R < 2$ and R is a number not constructible with straightedge and compass from the unit length and $\|r(1,0) - (1,0)\|$. An open dense subset S of C is eventually carried to the interior of the unit disc, and the endpoints of the arcs of S are carried to the circle. However, the complement T of S in C is homeomorphic to the Cantor set and its *non-endpoints* never reach the disc. For the proof, define

$$F_k = r^{-k} \circ f \circ r^k.$$

Then

$$(r \circ f)^n = r^{n+1} \circ F_n \circ F_{n-1} \circ \cdots \circ F_0.$$

Here F_j is the folding toward the disc along the tangent to $r^{-j}(1,0)$. Since r^{n+1} is an isometry, we may as well work with the sequence of compositions of planar foldings

G_k where

$$G_k = F_k \circ F_{k-1} \circ \cdots \circ F_0.$$

Let B denote the interior of the unit disc, S_n that part of C taken into B by G_n, and T_n the rest of C. Set

$$S = \cup S_n \subseteq C$$

and

$$T = C \cap S';$$

thus T is the complement of S in C. Clearly S is open in the relative topology of C. We need to show that S is dense in C and that T has no isolated points. The image under G_k of C is a piecewise circular curve. If $\overline{p_1 p_2}$ is a maximal arc in T_n, then for some $k > n$ we have

$$r^{-k}(1,0) \in \overline{G_n(p_1)G_n(p_2)} \quad \text{on } \mathrm{Bd}(B).$$

Hence $\overline{p_1 p_2}$ is not contained in T_k. Furthermore, $r^{-k}(1,0)$ is neither $G_n(p_1)$ nor $G_n(p_2)$ since otherwise R would be constructible. Consequently p_1 and p_2 are endpoints of arcs of positive length in T_j for all $j \geq n$, and for all n the set S_n is a finite union of open arcs which have no common endpoints. Thus, S is dense in C and T has no isolated points. Since any compact, perfect, totally disconnected set in R (or C) is homeomorphic to the Cantor set, we are done.

Editorial comment. This was solved also by the proposer, who indicated that certain generalizations to sequences of foldings in spaces of arbitrarily high dimension are forthcoming in a paper by J. Lawrence and J. Spingarn, "On iterates of nonexpansive mappings and foldings."

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

The Turing Omnibus: 61 *Excursions in Computer Science*. By A. K. Dewdney, Computer Science Press, W. H. Freeman and Company, New York, 1989. xiv + 415 pp.

MARK BRIDGER
Mathematics Department, Northeastern University, Boston, MA 02115

It is not clear exactly when people began referring to computer science as a "field" in the sense that mathematics, biology, and physics are fields. In fact, the question of whether it is a separate field at all is of some academic interest, since computer science professors have been housed in the engineering, mathematics, and physics departments of quite a few institutions. Many people—both within and without computer science—would say, when asked for a definition: "I don't know exactly what computer science is, but I know it when I see it (usually)."

A. K. Dewdney, the editor of the "Computer Recreations" column in *Scientific American*, has now provided us with a useful survey of some of the topics which he considers computer science, or at least those which computer scientists should know. *The Turing Omnibus* is aptly titled; it is a sight-seeing trip through computer country, including significant side excursions into some of the regions of computability theory analyzed by A. Turing (and others). In all, Dewdney has written 61 short essays, the average being 5–6 pages. These are divided into 11 groups: Analysis of Algorithms, Artificial Intelligence, Complexity Theory, Data Structures, Logic Design (hardware basics), Coding and Cryptography, Computer Graphics, Automata and Languages, Theory of Computation and Applications (CAT scans, linear programming, etc.). There is also a chapter of Miscellany which includes material on neural nets, perceptons, number systems, recursion, analog computers, and self-replicating machines.

It is impossible in this space to summarize each essay, but descriptions of some of the major groups will give the flavor.

In one excursion, Dewdney describes the four basic classes of automata (theoretical computing machines): finite, pushdown, linear bounded and Turing machines. He shows how their specifications impose structural conditions on their languages, leading to the four language types: regular, context free, context sensitive and recursively enumerable. He, thereafter, concentrates on the most powerful automata, the Turing machines, and uses them to look at various aspects of theoretical computability: deterministic polynomial time algorithms, NP and NP-complete problems, noncomputable functions (e.g., "busy beaver") and nonsolvable problems (e.g., halting).

In another byway we look at divide-and-conquer algorithms: mergesort, heapsort, fast multiplication, and fast Fourier transforms. There is also an essay on parallel processing. The algorithms are analyzed mathematically; for example, Dewdney shows that both mergesort and heapsort have speed $O(n \log(n))$, while parallel processing can reduce matrix multiplication from $O(n^3)$ to $O(\log(n))$.

The tour of Logic Design begins with a description of the basic "gates"—the simplest building blocks of computer circuitry. Later essays show how these are combined to form multiplexers, encoders and, ultimately, a "SCRAM": Simplified Complete Random Access Computer. The section ends with a quick tour of VLSI chip technology.

Computer graphics is somewhat underrepresented by essays on spline curves, Quad trees, and Sierpinski curves (this last is more a discussion of recursion than graphics). It would have been nice to see something here on ray-tracing, fractal imaging or hidden surface algorithms.

One interesting feature of the book is that the 11 tours are not presented as separate sections of the book, but each is arranged sequentially into subtours, which are then "shuffled" together. As the author puts it: "The mixing is deliberate. A chapter preparing in some important way for another will always precede it; but between the two, a sequential reader will discover a number of other chapters on quite different topics.... These show some of the crosslinks that actually exist in the subject." This idea must be termed a partial success. On one hand, the forced variety made the book easy to read—if one essay didn't strike my mood or interest, the next one often did. Sometimes it was hard to put the book down, since the effect was a kind of enforced browsing, made guilt-free since, after all, it was the author's intention.

On the other hand, Dewdney is not a very consistent writer. Some essays are written in an intuitive, informal style, while others are quite professorial and formal. For example, there are several essays centered around Nondeterministic Polynomial time (NP) problems. The first, dealing with the satisfiability problem, is quite understandable. However, the discussion of NP vs. NP-complete problems is confusing, and the reasons point up some problems with the author's style. First, the role of a nondeterministic machine in this context is simply to provide a guess which we assume to be correct (best possible case); thus, solving a problem means verifying a correct guess. Instead of just saying this, Dewdney writes: "If for all yes instances of the problem (and only for these) there is a guess which will cause the Turing machine to answer yes, then the program is said to solve the decision problem. It does so in polynomial time if there is some fixed polynomial that bounds the number of steps in at least one yes computation for each possible yes instance the Turing machine might encounter. In such a case, the problem is said to lie in NP." While this is technically true, it is somewhat obscure, and certainly difficult to read. The problem is compounded by the author's tendency to use terms before they have been defined; in fact, the chapter on NP-completeness does not contain a definition for the term, though hints are given. It is only in later essays that we may find out that problems are NP-complete if the deterministic solution of any *one* of them in polynomial time implies that *every* problem in NP is (deterministically) solvable in polynomial time. The fact that these later essays are scattered throughout the remainder of the book just makes putting the pieces together more difficult.

The reader should also note that there are some confusing typos; nothing serious, just don't believe that every circuit is exactly correct as drawn, or that every checker move is legal (!). Dewdney also seems to miss the role that second derivatives play in splines, and that solving simultaneously for *all* spline coefficients is, in fact, $O(n)$ (where n is the number of points to be fitted). Also, the essay on CAT scanning is

just a summary of the abbreviated account in Pavlidis' book *Algorithms for Graphics and Image Processing*. Nevertheless, there are many more vistas—such as the excellent 6-page condensation of Gödel's proof of the Incompleteness Theorem—than (permit me) billboards.

At one time Dewdney intended this book to be accessible to the "educated layperson"; he even included a section on algorithms-as-recipes. Nevertheless, things turned out differently, and a reader without some training in theoretical science would have a difficult time indeed with most of these essays. However, technically oriented people, especially mathematicians, computer science students and professionals, will have a challenging and rewarding time surveying the landscape presented, and trying some of the inspirational exercises. Travel broadens, it is said, and nearly anyone will learn something from *The Turing Omnibus*.

Upon finishing the book it occurred to me that 25 years ago many of the topics Dewdney has included would have been considered exclusively in the province of mathematics or philosophy. These include Boolean algebra, graph theory, splines (as part of curve fitting), Fourier transforms (fast and otherwise) and the predicate calculus. In fact, the whole methodology of modern abstract computer science has been borrowed from the abstractions of formal logic and, more generally, theoretical mathematics. However, computer science is no more just mathematics than, say, physics is. Reading papers in modern theoretical physics is a lot like reading mathematics, but that doesn't make physics a branch of mathematics. Physics starts with an attempt to model something given: the real world; this informs and shapes the mathematics used and created by the physicist. Similarly, computer science is ultimately bound to the questions: What problems can a computing machine solve, and how can the machine be designed to solve them efficiently? The flavor of the mathematics that is used to answer these questions is quite a bit different from that constructed by the "pure" mathematician or even the theoretical physicist. Finiteness, computability, and efficiency are the criteria which govern the direction of computer science research, and developments in this new field have profoundly influenced some of the directions taken in its more mature relatives.

A Course in Constructive Algebra. By R. Mines, F. Richman, and W. Ruitenberg, Springer-Verlag Universitext, New York, 1988. xi + 344 pp.

MICHAEL J. BEESON
Department of Mathematics and Computer Science, San Jose State University, San Jose, CA 95192

What is Constructive Algebra?. Elementary algebra is full of algorithms: factoring polynomials, solving equations, inverting matrices. Advanced algebra is full of non-constructive arguments, for example "constructing" the algebraic closure of a field by Zorn's lemma. The question arises: to what extent can algebra be made constructive? If we are careful, can we eliminate things like Zorn's lemma that don't correspond to algorithmic constructions, or are they essential to the mathematical theory? The book under review contains the answers.

The word "constructive" is not always used in the same way by different mathematicians. I shall therefore begin by explaining what I mean by it in this review. First, by an *algorithm* I understand a procedure that could be programmed for a computer, but without regard to questions of efficiency. For example, Kronecker gave algorithms in 1881 for factoring polynomials in several variables

over the integers. These algorithms are hopelessly inefficient, and much better algorithms are known today; but they *do* nevertheless count as algorithms. Second, by a "correctness proof" of an algorithm, I mean a proof that if the input data satisfy certain conditions, then the algorithm terminates (produces output) and the output data satisfy certain conditions. For example, a factoring algorithm accepts as input data a polynomial with integer coefficients, and terminates producing a list of polynomials, satisfying the condition that their product is the input polynomial.

With the concept of an algorithm understood, we may call a piece of mathematics *constructive* if it consists of

(1) definitions of certain data structures (e.g., polynomials over the integers)
(2) definitions of certain algorithms over those data structures
(3) correctness proofs of those algorithms.

The above cannot be taken as a *definition* of "constructive", because it would then be circular; the correctness proofs in (3) have to be constructive themselves. Rather, it is intended as an aid to the intuition. The concept of "constructive proof" is both as simple and as difficult to define as the usual concept of "correct proof." It is not usual to begin teaching mathematics by *defining* the notion of "proof"; nor can that really be done, as we have no more primitive concept in terms of which to couch the definition. The same goes for "constructive proof." About such fundamental philosophical matters, we shall say no more in this review.

Constructive algebra versus ordinary algebra. Until the early twentieth century, the predominant view of most mathematicians was that mathematics consists of the three steps outlined above as "constructive mathematics." The work of Hilbert broke new ground by making extensive use of existence proofs by contradiction, which nobody knew how to fill in by algorithms. Similarly, Zermelo's non-constructive proof of the existence of a well-ordering of the reals created a stir in 1906. The widespread acceptance of existence proofs not accompanied by algorithms did not really begin until about 1930; so we only have a half-century of non-constructive mathematics. Nevertheless, it has been a very productive half-century, so the effort to consolidate its gains from the constructive viewpoint must be considerable.

This consolidation was begun by Errett Bishop in his 1967 book, *Foundations of Constructive Analysis*. He deliberately did not attempt to constructivize modern algebra. Since 1967, then, algebra has been crying out for a systematic constructive treatment like that given to its sister. The book under review is a sister book to Bishop's, doing for algebra what Bishop did for analysis. It differs from Bishop's book in one way: there has been a deliberate attempt to make it usable as a textbook.

The authors are to be congratulated for their success in providing an algebraic companion to Bishop's volume. No doubt Bishop would be pleased with their work and very much appreciate the dedication to him.

Constructive algebra versus algorithmic algebra. Today nearly every mathematician has at least heard of computer programs that carry out symbolic computation, such as MACSYMA, Mathematica, MAPLE, CAYLEY, and SCRATCHPAD. These programs embody steps (1) and (2) above for certain predefined data structures and algorithms, and to some extent (particularly in SCRATCHPAD) they allow the user to define new data structures and algorithms. The development and use of such programs has awakened new interest in constructive mathematics. This

turn in the way mathematics is viewed is not really a new direction, but a return to tradition, as discussed above, although on the time scale of one human life it may appear to be a new direction.

By "algorithmic mathematics," I mean something similar to constructive mathematics, but with a difference: in algorithmic mathematics, we are concerned with the *actual implementation* of the algorithms. Thus Kronecker's factorization algorithms are constructive, but not of interest for algorithmic mathematics; for actual implementation, we need the modern factorization algorithms (see Knuth vol. 2 for an introduction). The number of people who are actively using computerized symbolic computation is in the thousands; and several hundreds are involved in research. By contrast, the community of people who have read and understood the work of Errett Bishop probably numbers not more than a few hundred, while about ten are involved in research.

I think it is safe to say, moreover, that the intersection of these two groups is very small. The number of people who could give you a coherent explanation of Bishop's definition of real numbers and also describe what Buchberger did about Gröbner bases[1] (this is fundamental work in algorithmic algebra) is vanishingly small. (Of course it's impossible to know the exact number, but I could not name ten such people.) This is, in my view, an unfortunate situation.

There would be benefits all around from more communication. The followers of Bishop ought to read the literature in symbolic computation. The algorithmic algebraists could benefit considerably from reading both Bishop and the volume under review.[2] The classical algebraists should read both, and play with the symbolic manipulation programs. Everybody should listen more and argue less.

The book under review appears to have been intended, not for the mathematicians involved with algorithmic algebra, but for the ordinary mathematician, perhaps while teaching a course in algebra. The authors hope that such mathematicians will consult the book while preparing their lectures, or even use it as a textbook. It seems to be well-written for such a purpose.

On the other hand, in 1988 twenty-one years had passed since Bishop's book, and during those twenty-one years the "computerization of mathematics" has begun. Bishop himself was interested in the actual computerization of mathematics. I do not think it an unfair criticism to say that in 1988, it would have been appropriate to take the efficiency of algorithms into account while writing a volume on constructive algebra. If the algorithmic algebraists *do* read the volume under review, they will not find a systematic development of the mathematics that underlies their algorithms, but instead a systematic development of a constructive-in-principle-but-

[1]A Gröbner basis for a polynomial ideal is a finite set of polynomials g_i such that the rules $f \to f \bmod g_i$ form a terminating, confluent set of rewrite rules, i.e., if we apply these rules in any order repeatedly to a polynomial f, we always arrive at a unique stopping point. What Buchberger did was give an algorithm for the construction of a Gröbner basis. Once we know how to construct a Gröbner basis, many other questions about polynomial ideals have easy algorithmic solutions. See Buchberger and Loos (1982), p. 37; and trace the bibliographical references in that volume for more information on algorithmic algebra.

[2]I understand there is a forthcoming book by J. Davenport on the subject of algorithmic algebra; I do not know its title or projected date of publication, but it should be a good companion to the book under review.

not-in-practice kind of algebra. I maintain that this would still be a good and useful thing for them to read; whether they will perceive the matter in this way, I doubt.

The flavor of the subject. Now let me turn to a more detailed consideration of the contents of the book. My aim is not to outline the book, but rather to provide a few examples of the kind of considerations that arise when one does algebra constructively.

The first distinction to be made is that between *discrete* structures and not necessarily discrete structures. A set is called *discrete* if we have an algorithm for deciding equality on the set. This definition presumes the constructive viewpoint in which one manipulates data structures representing elements of an abstract set. For example, real numbers are given by programs for generating arbitrarily close rational approximations. The reals are an example of a non-discrete structure: there is no algorithm for deciding if a program represents the real number zero or not.[3] So far, in algorithmic algebra only discrete structures have been used; and for good reason, as the constructive theory of non-discrete algebraic structures is fraught with difficulties. In order not to obscure the overall picture, the authors have chosen to treat primarily the case of discrete structures. This was a wise choice: the subtleties of the non-discrete theory would have exhausted the patience of most readers.[4] Among the discrete structures are all finite structures, the rationals, polynomial rings over discrete coefficient rings, and the algebraic numbers. Among the non-discrete structures are the reals and the complex numbers.

In addition, the systematic theory must be prepared to cope with a structure like the field $\mathbf{R}(ia)$, obtained from the reals by the adjunction of a purely imaginary number ia. The difficulty here is that we have no algorithm to determine (as a function of the real number a) whether this field is equal to $\mathbf{R}$ or to $\mathbf{C}$, since we can't tell if a is zero or not. Such fields do not arise in algorithmic mathematics, which has not yet considered algorithms whose input and output are algebraic structures. All the algorithms so far implemented deal only with fixed groups, rings, and fields, and manipulate their elements only.[5]

Now, for example, the Euclidean algorithm for the greatest common divisor (perhaps the most fundamental algorithm in algorithmic algebra) finds its proper context in Theorem 5.6 of the book: it works in a discrete commutative ring with "recognizable units." This means that there is an algorithm for testing whether an element is a unit or not.

The next most fundamental algorithm is factorization. On p. 109 of the book we find the fundamental definition (here rephrased): an element p is *irreducible* if we have an algorithm to determine (given input a and b) that one of a or b is a unit, and which one. A UFD is then defined as a discrete domain R with a factorization algorithm, i.e., an algorithm for determining of a nonzero element r either that r is

[3]Computer implementations of "real numbers" are usually either of rational numbers with a fixed bound on the denominator ("floating point") or of rational numbers without a fixed bound on the denominator ("bigfloats"). Computer mathematics has yet to catch up to Weierstrass.

[4]An example of the complications: In the non-discrete case we want to consider, e.g., groups with an apartness relation. Under what circumstances can we carry over an apartness relation to a factor group? For example, there is no natural apartness on $\mathbf{R}/\mathbf{Q}$, the reals mod the rationals.

[5]Algorithmic mathematics has yet to catch up to Noether and Hilbert, as well as Weierstrass.

a unit, or for producing a factorization of r into irreducibles.[6] The famous theorems of Kronecker (1881) find their statements as follows: Let R be an infinite UFD with finitely many units. Then $R[x]$ is a UFD. Moreover, if $R[x]$ is a UFD, so is $R[x, y]$. That is, factorization of elements implies factorization of polynomials, even in several variables. In algorithmic algebra, we would go on to look for efficient factorization algorithms. In constructive algebra, we go on instead to look at the dependence on the data types: can we remove the hypothesis that the ring R be infinite? After all, in the case of a finite ring, there are only finitely many divisors of a given polynomial, so in principle we could examine them all; hence in case R is finite we are done. But we don't have an algorithm to determine from the presentation of a ring (e.g., as a factor ring) whether it is finite or not; so we can't make the case distinction between the search algorithm (in the finite case) and Kronecker's algorithm (in the infinite case). (See p. 126 of the book.)

In Chapter V of the book, there is a nice treatment of the decomposition theorem for modules over a PID, with applications to Jordan canonical form and later, in Chapter VI, to Galois theory. In Chapter VI, there is a construction of a splitting field for a nonzero polynomial over a countable discrete field. The construction of splitting fields is a tricky business, since we may not be able to factor polynomials. If f is a polynomial, then $k[X]/(f)$ is only a *ring*; it is a *field* only if f is irreducible, and the classical technique of working with an irreducible factor of f is not constructive, since we may not have an algorithm for finding such a factor. Instead, we consider $k[X]/M$, where M is a maximal ideal containing (f), constructed using an enumeration of the base field k (bottom of page 151 in the book). Without the assumption that k is countable (and the enumeration algorithm implicit in that assumption) we can't construct a field in which f has a root. This construction does not support a proof of uniqueness of the splitting field, and indeed Exercise 4 (p. 153) shows that uniqueness is not constructively true.

Similarly, the construction of an algebraic closure of a discrete field k relies on countability. The classical existence proof of an algebraic closure of a field k is highly non-constructive, appealing as it does to Zorn's lemma or a transfinite form of the axiom of choice. There are interesting cases still open, between the theorem just mentioned and the classical but non-constructive proof. For example, what about the field $\mathbf{C}(t)$ of rational functions with complex coefficients? Can you construct its algebraic closure? (If so, write the reviewer.)

Questions about the zeroes of polynomials have a central place in algebra, starting with the fundamental theorem of algebra. A number of constructive proofs of the fundamental theorem of algebra are known (see the Historical Appendix of Beeson (1985)). A purely algebraic proof, using Galois theory and the existence of 2-Sylow subgroups of finite groups, is given in the book.

[6] I have phrased these definitions in terms of algorithms to avoid the usual squabbling about the meaning of "there exists." The authors, like Bishop and his predecessors back to Brouwer, use "there exists" as a shorthand for assertions about algorithms which are implicit in the discussion. Other people seem unable to accept this redefinition of the phrase "there exists," even for the moment by "willing suspension of disbelief," and instead of considering the mathematics, argue about the phrase. Worse, the constructivists argue back! claiming, for example, that it isn't a redefinition but a definition, the phrase never having been properly defined. In this review I shall avoid the whole linguistic problem by speaking directly of the algorithms instead.

An interesting result about factorization is this: Call a field k *fully factorial* if we have a polynomial factorization algorithm for every finite extension $k(\alpha_1, \ldots, \alpha_n)$. (Here the α_i are inputs to the algorithm.) Can we characterize the fully factorial fields? It turns out that we can: the right condition is "Seidenberg's condition P": there must be an algorithm for deciding whether a finite set of linear equations over k is solvable in k^p (the set of pth powers in k), and finding the solution if it exists.

Condition P is closely connected to Wedderburn's theorem as well, as we find out several chapters further on (p. 244): a discrete field satisfies condition P if and only if each finite-dimensional k-algebra has a finite-dimensional radical. That is, if we have an algorithm for passing from a basis for the k-algebra to a basis for the radical. The book goes on to give a constructive version of Wedderburn's structure theorem for semi-simple algebras. The next chapter gives a constructive treatment of Ulam's theorem on the structure of Abelian groups.

All the chapters from Jordan canonical form onwards have this feature: the main difficulties arise because of the dependence of algorithms on the (presentation of the) ring or field, rather than on the elements. The algorithms involved, to the extent that they go beyond the usual undergraduate curriculum, are unusual in that they make use of the hypothesized presentation of the field, as in the constructions of splitting fields or algebraic closures described above. In the last chapter, on valuation theory, the algorithmic algebraist will feel at home again. Here we meet our old friend, Hensel's lemma, and the wonderful corollaries of the algorithm embodied in Hensel's lemma. As it happens, the efficient modern polynomial factorization algorithms depend on Hensel's lemma and on factorization in finite fields, using the GCD algorithm in the finite field. No mention is made in the book of this amazing connection between the last chapter and the early chapters (for which see Knuth vol. 2), but the systematic constructive theory of valued fields is nicely developed.

REFERENCES

1. M. Beeson, Foundations of Constructive Mathematics, Springer-Verlag, Berlin/Heidelberg/New York, 1985.
2. E. Bishop, Foundations of Constructive Analysis, McGraw-Hill, New York, 1967. Second edition co-authored by D. Bridges, Springer-Verlag, Berlin/Heidelberg/New York/Tokyo, 1985.
3. B. Buchberger and R. Loos, Algebraic simplification in: Computer Algebra: Symbolic and Algebraic Computation, edited by Buchburger *et. al.*, pp. 11–44, Springer-Verlag, Wien/New York, 1982.
4. D. E. Knuth, Art of Computer Programming, vol. 2, Seminumerical Algorithms, Addison-Wesley, Reading, Mass., 1981.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, P. *Eight Papers Translated from the Russian.* S.G. Dalalyan, *et al.* AMS Transl., Ser. 2, V. 145. AMS, 1989, vii + 117 pp, $54. [ISBN: 0-8218-3129-1]

Precalculus, T(13: 1). *College Algebra, Second Edition.* Jerome E. Kaufmann. PWS-Kent, 1990, xiv + 562 pp, $32. [ISBN: 0-534-92095-0] College algebra text intended for precalculus, pre-finite mathematics, or terminal course. Spiral sequencing: e.g., quadratic equations and parabolas are presented in each of chapters 2 (equations, inequalities, and problem solving), 4 (functions), and 9 (conic sections). Mostly rote or manipulative problems. FR

Precalculus, T(13). *College Algebra and Trigonometry, Second Edition.* Jerome E. Kaufmann. PWS-Kent, 1990, xiv + 737 pp, $33. [ISBN: 0-534-92104-3] The minor changes in this edition include a few rewritten sections and added coverage of scientific notation, piecewise-defined and even and odd functions. (*First Edition,* TR, June-July 1987.) JNC

Precalculus, T. *Precalculus Mathematics: A Functional Approach, Fourth Edition.* Karl J. Smith. Brooks/Cole, 1990, xvii + 602 pp, $44.75. [ISBN: 0-534-11922-0] This is the *Fourth Edition* of a successful precalculus text. (TR, *Third Edition,* November 1986.) WE

History, P*, L. *The History of Ptolemy's Star Catalogue.* Gerd Grasshoff. Studies in the History of Math. & Physical Sci., V. 14. Springer-Verlag, 1990, xiv + 347 pp, $63. [ISBN: 0-387-97181-5] Covers the time in which the Almagest defined the standards of astronomy and later. Cites difficulties Arabic astronomers had with the star catalogue. Examines the period in which the accusation of forgery by Ptolemy was raised. Summarizes the efforts of twentieth-century historians to rehabilitate Ptolemy as an accomplished observer. Earlier chapters on work of others. Later chapters comprise a critical evaluation of the earlier ones. Numerous footnotes and references. Appendices are laden with tables and charts. JDEK

Logic, T(17-18: 1). *Computability and Logic, Third Edition.* George S. Boolos, Richard C. Jeffrey. Cambridge U Pr, 1989, xi + 304 pp, $49.50; $17.95 (P). [ISBN: 0-521-38026-X; 0-521-38923-2] New chapters on undecidable sentences and nonexistence of nonstandard recursive models of elementary Peano arithmetic. Some new exercises and revisions in other chapters. (*First Edition,* TR, March 1975; *Second Edition,* TR, November 1981.) KES

Foundations, P. *Lecture Notes in Mathematics-1401: Set Theory and Its Applications.* Eds: J. Steprāns, S. Watson. Springer-Verlag, 1989, 227 pp, $20 (P). [ISBN: 0-387-51730-8] The proceedings of a conference at York University, Toronto, Canada, August 10-21, 1987, attended by 80 mathematicians from 12 countries. Most applications are to general topology. LCL

Graph Theory, P, L. *Every Planar Map is Four Colorable.* Kenneth Appel, Wolfgang Haken. Contemp. Math., V. 98. AMS, 1989, xv + 741 pp, $80 (P). [ISBN: 0-8218-5103-9] Everything you need to know about the famous Appel-Haken proof of the Four Color Conjecture in one quite thick volume: an expository introduction for non-specialists; emended reprints of the two original 1977 papers from the *Illinois Journal of Mathematics* that contain the proof; a newly-written Appendix that fills in important details and proves that the algorithm for map coloring is polynomial-time; and a copy on paper of the 500-page supplement (originally provided on microfiche) that provides details of reduction lemmas and class check lists. The original proof, as critics at the time claimed, was not without error; but the authors too were right: all errors so far discovered are correctable, and have been dealt with in this twelfth anniversary compendium. LAS

Combinatorics, T(17-18: 2), P, L. *Combinatorics of Finite Sets.* Ian Anderson. Oxford U Pr, 1987, xv + 250 pp, $47.50. [ISBN: 0-19-853367-5] Studies properties of collections of subsets of a finite set, where the collections are described in terms of intersection, union, or inclusion conditions. Sperner's theorem is the starting point of the discussion. Chapter topics include symmetric chains, the Erdös-Ko-

Rado theorem, the Kruskal-Katona theorem, generalized Macaulay theorem for multisets, and the Littlewood-Offord problem. Exercises, hints, and solutions. (Extended Review, May 1989.) LC

Combinatorics, P. *Invariant Theory and Tableaux.* Ed: Dennis Stanton. Inst. for Math. & its Applic., V. 19. Springer-Verlag, 1990, xi + 298 pp, $36. [ISBN: 0-387-97170-X] Proceedings of the workshop on invariant theory and tableaux held at the Institute for Mathematics and Its Applications, Minneapolis, Minnesota in March 1988. The papers have been roughly grouped into four parts: invariant theory, Coxeter groups, enumeration, and tableaux. LC

Combinatorics, T(16-18), S, P*, L*. *The Stable Marriage Problem: Structure and Algorithms.* Dan Gusfield, Robert W. Irving. Foundat. of Comput. Ser. MIT Pr, 1989, xvii + 240 pp, $27.50. [ISBN: 0-262-07118-5] A coherent exposition of most of the algorithmic and structural results that are known about stable matching problems. A special feature is the emphasis on a more general and algebraic context ("rings of sets"), which leads to a clearer understanding of the relationships between a number of variants (stable marriage problem, stable roommates problem, etc.), and to improvements in the design of efficient algorithms. The book concludes with a discussion of twelve open problems, and a large bibliography. This very active area of research shows no sign of drying up! LCL

Discrete Mathematics, T*(13-14: 1), L. *Discrete Mathematics with Algorithms.* Michael O. Albertson, Joan P. Hutchinson. Wiley, 1988, xiii + 546 pp, $41. [ISBN: 0-471-84902-2] Text for computer science or mathematics students. Algorithmic reasoning is unifying theme. Treats proofs of correctness and efficiency analysis. Covers sets, functions, proofs by induction and contradiction, permutations and combinations, number theory, graph theory, searching and sorting, recurrence relations. No programming experience or calculus assumed. KES

Number Theory, P, L. *Modular Forms.* Toshitsune Miyake. Springer-Verlag, 1989, ix + 335 pp, $73. [ISBN: 0-387-50268-8] Translation of part of the Japanese "Automorphic forms and number theory" (1976), introducing the theory of modular forms of one variable. Includes chapters on Fuchsian groups, automorphic forms and Hecke algebras, L-functions, modular groups and modular forms, unit groups of quaternion algebras, traces of Hecke operators, and Eisenstein series. With up-to-date references and extensive numerical tables. SB

Number Theory, P. *Lecture Notes in Mathematics-1406: Analytic Theory of Continued Fractions III.* Ed: L. Jacobsen. Springer-Verlag, 1989, 142 pp, $13.50 (P). [ISBN: 0-387-51830-4] Proceedings of a research seminar-workshop held in Redstone, Colorado from June 26 to July 5, 1988. Contains ten papers, including some survey articles and some current research. RH

Number Theory, P. *Algorithms for Diophantine Equations.* B.M.M. de Weger. CWI Tract, V. 65. Mathematisch Centrum, 1989, vii + 212 pp, Dfl. 33 (P). [ISBN: 90-6196-375-3] A monograph on practically effective algorithms for solving Diophantine equations (as opposed to theoretically effective algorithms). Starting with Gelfond-Baker theory of linear forms in logarithms, the author develops methods to go from bounds on size of solutions to actual solutions. Includes specific examples. GG

Number Theory, T*(17-18), S*, P*, L*. *Modular Functions and Dirichlet Series in Number Theory, Second Edition.* Tom M. Apostol. Grad. Texts in Math., V. 41. Springer-Verlag, 1990, x + 204 pp, $49.80. [ISBN: 0-387-97127-0] Designed as the second volume to the author's *Introduction to Analytic Number Theory* (Springer UTM; TR, November 1976). Devoted primarily to elliptic and modular functions and their number theoretic applications. Changes in this edition include an alternate treatment of the transformation formula for the Dedekind eta function (supplement to Chapter 3), misprint corrections, updated bibliography, and minor exercise changes. Assumes background in number theory comparable to the first volume and basic concepts of complex analysis. (*First Edition,* TR, January 1977.) SB

Linear Algebra, T*(15-16: 1). *Linear Algebra.* Bill Jacob. WH Freeman, 1990, xvii + 547 pp, $37.95. [ISBN: 0-7167-2031-0] Part I contains the usual content of an introductory semester course with a gradual increase in the level of abstraction. Part II covers bilinear forms, symmetric matrices, quadratic forms, structure of linear operators, applications to infinite-dimensional vector spaces, and multilinear algebra. A clear and concise presentation. JNC

Group Theory, T(18: 1, 2), S, P. *Combinatorial Group Theory: A Topological Approach.* Daniel E. Cohen. Math. Soc. Stud. Texts, V. 14. Cambridge University Pr, 1989, x + 310 pp, $49.50; $19.95 (P). [ISBN: 0-521-34133-7; 0-521-34936-2] Can be used as a one-semester course on either group theory or topology, but the real theme of this book is the interplay between these two branches of mathematics. An expanded version of notes produced about ten years ago in the Queen Mary College Lecture Note Series. Includes a full account of Bass-Serre theory and discusses the unsolvability of the word problem. LCL

Group Theory, S(18), P. *Lecture Notes in Mathematics-1398: Groups—Korea 1988.* Eds: A.C. Kim, B.H. Neumann. Springer-Verlag, 1989, 189 pp, $16.20 (P). [ISBN: 0-387-51695-6] Proceedings of a conference on group theory held in Pusan, Korea, August 1988. Includes twenty complete papers, lists of one-hour and seminar talks, and a list of participants. JS

Algebra, S(18), P. *Primes Associated to an Ideal.* Stephen McAdam. Contemp. Math., V. 102. AMS, 1989, xiii + 167 pp, $31 (P). [ISBN: 0-8218-5108-X] Discusses five sets of prime ideals associated to an ideal I in a Noetherian ring: the persistent, asymptotic, quintasymptotic, essential, and quintessential

primes of I. The aim is to present most of the important results in one book. Accessible to graduate students with background in commutative algebra. Intended mainly for commutative ring theorists, but may also be of interest to noncommutative ring theorists and algebraic geometers. RH

Algebra, S(16-17), P*, L*. *Algebra I: Basic Notions of Algebra.* Eds: A.I. Kostrikin, I.R. Shafarevich. Encycl. of Math. Sci., V. 11. Springer-Verlag, 1990, 258 pp, $59.80. [ISBN: 0-387-17006-5] A marvelous survey of modern algebra by one of its masters, I.R. Shafarevich. The reader obtains a sense of the scope and depth of algebra and a real feeling for the algebraic point of view in mathematics. A book that any mathematician will value. SG

Algebra, S(18), P*. *Cyclotomic Fields I and II, Combined Second Edition.* Serge Lang. Grad. Texts in Math., V. 121. Springer-Verlag, 1990, xvii + 433 pp, $49.50. [ISBN: 0-387-96671-4] Combines the two previous volumes (with corrections). This edition includes a purely cyclotomic proof of "the main conjecture" (Theorem of Mazur-Wiles). (*Volume I,* TR, April 1979; *Volume II,* TR, December 1980.) SB

Calculus, T*(13: 1). *Calculus for Management, Social and Life Sciences, Second Edition.* Dennis D. Berkey. Saunders College, 1990, xvi + 469 pp, $36.50. [ISBN: 0-03-031264-7] Additions to this edition include new examples and exercises, a new chapter on trigonometric functions, and expanded treatment of logarithms and elasticity of demand. Illustrations are now computer-generated. (*First Edition,* TR, October 1986.) JNC

Real Analysis, T(18: 1), S, P. *Lecture Notes in Mathematics-1381: Weighted Hardy Spaces.* Jan-Olov Strömberg, Alberto Torchinsky. Springer-Verlag, 1989, v + 193 pp, $16.20 (P). [ISBN: 0-387-51402-3] An advanced textbook which includes results on Hardy spaces from within the past ten years. Authors are among several contributing to subjects' development, but their book takes broad view. One unusual stylistic feature: statements of lemmas and theorems are often placed together, to be followed by a succession of proofs. GG

Real Analysis. *Almost Periodic Functions.* C. Corduneanu, N. Gheorghiu, V. Barbu. Transl: Gitta Berstein, Eugene Tomer. Chelsea, 1989, x + 257 pp, $23.95. [ISBN: 0-8284-0331-7] Almost periodic functions are (as defined here) uniform limits of trigonometric polynomials; hence they admit suitable versions of harmonic analysis, differential equations methods, etc. This slightly revised edition of the first English edition, published in 1968, covers elementary theory and applications (TR, June-July 1969). Style is leisurely, readable, and essentially self-contained, assuming only beginning graduate-level analysis. Includes a colossal bibliography, updated to around 1987. PZ

Functional Analysis, T(16-17: 1), L. *Analyse Fonctionnelle.* M. Samuelides, L. Touzillier. Cepadues-Editions, 1989, 289 pp. [ISBN: 2-85428-233-7] A concise and careful introduction intended for engineers. Includes exercises and a chapter on the history. JD-B

Analysis, P. *Orthogonal Polynomials and Their Applications.* Ed: Jaime Vinuesa. Lect. Notes in Pure & Appl. Math., V. 117. Marcel Dekker, 1989, x + 207 pp, $99.75 (P). [ISBN: 0-8247-8161-9] Proceedings of the International Congress on Orthogonal Polynomials held in September 1987 at Laredo, Spain. LC

Analysis, T(17: 1), S. *Point Set Theory.* John C. Morgan II. Pure & Appl. Math., V. 131. Marcel Dekker, 1990, viii + 279 pp, $99.75. [ISBN: 0-8247-8178-3] A unified treatment of Baire theory, Lebesgue theory, and point set topology which requires a background in all three for full appreciation. Written in concise definition-theorem-proof-example style. Contains extensive bibliography, but no exercises. Over 50% discount on price for classroom use. GG

Algebraic Geometry, P. *Lecture Notes in Mathematics-1399: Arithmetic of Complex Manifolds.* Eds: W.-P. Barth, H. Lange. Springer-Verlag, 1989, 170 pp, $16.20 (P). [ISBN: 0-387-51729-4] A collection of twelve papers presented at a 1988 conference in Erlangen. SG

Geometry, P. *Recent Developments in Geometry.* Eds: S.-Y. Cheng, H. Choi, Robert E. Greene. Contemp. Math., V. 101. AMS, 1989, xiii + 338 pp, $39 (P). [ISBN: 0-8218-5107-1] Out-growth of the special session on geometry at the AMS meeting at UCLA in November 1987. Purpose is to provide geometers with insight into recent developments in a variety of different specialties in geometry. Papers are surveys of relatively broad areas of geometry, not presentations of new results as such. JDEK

Operations Research, S(15-16), C, P. *Exploring Operations Research and Statistics in the Micro Lab.* Erwin Kalvelagen, Henk C. Tijms. Prentice-Hall, 1990, ix + 192 pp. [ISBN: 0-13-297334-0] The textbook and accompanying software package (for an IBM PC or compatible) is designed to be used as a laboratory supplement to courses in operations research and statistics. Emphasizes model formulation and the interpretation of computer-generated solutions. Topic areas include linear and integer programming, dynamic programming, and probability and statistics (the law of large numbers, simulation, and the central limit theorem). AO

Operations Research, P, L. *Knowledge-Based Process Planning for Construction and Manufacturing.* Carlos Zozaya-Gorostiza, Chris Hendrickson, Daniel R. Rehak. Academic Pr, 1989, xi + 306 pp, $39.95. [ISBN: 0-12-781900-2] A monograph describing the knowledge-based process planner PLANEX, a software system capable of aiding the planning of very diverse projects. While the authors focus on the construction of buildings, the system is flexible enough to handle bridges, highways and, as illustrated with an example, manufacturing operations for automobile electrical wire harnesses.

The book is intended for two audiences. First, developers of practical process planning tools, and second, researchers of planning systems. SM

Operations Research, T(15-16), S, L. *Mathematical Modelling: A Case Study Approach.* Dick Clements. Cambridge U Pr, 1989, viii + 166 pp, $44.50. [ISBN: 0-521-34340-2] A two-part textbook of mathematical modelling. Part 1 consists of a brief description of the modelling process, pitfalls to be aware of in applying mathematical methods to practical problems, and a chapter on pedagogical concerns arising from a modelling course focused on case studies. Part 2 contains the seven case studies presented in the form of internal corporate memos and blueprints. After each collection of memoranda, one possible solution to the problem is presented. SM

Dynamical Systems, P. *From Data to Model.* Ed: Jan C. Willems. Springer-Verlag, 1989, 246 pp, $59. [ISBN: 0-387-51571-2] Five papers cover various aspects of model selection and estimation procedures for time series data. RWJ

Control Theory, P. *Lecture Notes in Control and Information Sciences-131: Control of Large Flexible Space Structures.* S.M. Joshi. Springer-Verlag, 1989, ix + 196 pp, $30.30 (P). [ISBN: 0-387-51467-8] This research monograph discusses the unique challenges of controlling the very large and very fragile structures which will be constructed in earth orbit in the reasonably near future. The light weight and expansiveness of the structures will lead to extremely low frequency elastic modes. At the same time, highly accurate attitude controls are required. The book focuses primarily on the author's research into these questions, with extensive references to the work of others. Indexed. SM

Stochastic Processes, P. *Asymptotic Methods in the Theory of Stochastic Differential Equations.* A.V. Skorokhod. Transl. of Math. Mono., V. 78. AMS, 1989, xvi + 339 pp, $119. [ISBN: 0-8218-4531-4] This book, translated from Russian, was the idea of the author and Iosif Il'ich Gikhman, who died before he could take part in the writing. It covers asymptotic problems for stochastic differential equations, including ergodic theory for Markov processes, solutions for systems of stochastic differential equations, stability theory, averaging, and linear systems of equations. The coverage is very theoretical. Note the price. JJ

Statistics, S(17), P, L. *Rank Tests with Estimated Scores and Their Application.* Konrad Behnen, Georg Neuhaus. BG Teubner Stuttgart, 1989, ix + 416 pp, DM 54 (P). [ISBN: 3-519-02728-3] Develops a new class of nonlinear rank tests to compare how two distributions differ. Includes location, scale, symmetric difference, and independence tests. RWJ

Statistics, P. *Group Invariance Applications in Statistics.* Morris L. Eaton. Reg. Conf. Ser. in Prob. & Stat., V. 1. Institute of Mathematical Statistics (3401 Investment Blvd., No. 7, Hayward, CA 94545), 1989, v + 133 pp, $25 (P). [ISBN: 0-940600-15-3] Lectures given at an NSF/CBMS Regional Conference at the University of Michigan in June 1987. RSK

Statistics, S(17-18), P. *Methoden zur Schätzung der Ordnung bei autoregressiven Modellen.* Hartmut Maier. Teubner-Texte zur Math., B. 113. BG Teubner Leipzig, 1989, 148 pp, DM 18,50 (P). [ISBN: 3-322-00737-5]

Statistics, T(18: 1), S. *Model Discrimination for Nonlinear Regression Models.* Dale S. Borowiak. Statistics: Textbooks & Monographs, V. 101. Marcel Dekker, 1989, xi + 177 pp, $89.75. [ISBN: 0-8247-8053-1] Provides the practitioner with techniques and the statistician with basic theory in the area of model discrimination for nonlinear regression models. Assumes a set of competing nonlinear regression models have been established for explaining empirical data. Diagnostic statistics called discrimination functions are constructed and used on the empirical data to help in selecting the best model. Prediction ability of the model and parameter stability are both considered. Models looked at from the view of social scientists and engineers. When initial model assumptions of normality are violated resampling techniques, such as jackknife, bootstrap, and simulation, are used. Basic knowledge of statistical theory (estimation, distributional theory, hypothesis testing) required. Applications can be used without understanding proofs. JJ

Statistics, P. *Handbook of Nonlinear Regression Models.* David A. Ratkowsky. Stat.: Textbooks & Monographs, V. 107. Marcel Dekker, 1990, ix + 241 pp, $69.75. [ISBN: 0-8247-8189-9] Divided into two parts. Part I gives background to and philosophy of nonlinear regression modelling. Part II presents the most commonly used nonlinear regression models containing at most six parameters, together with the statistical properties of the estimators of the parameters. Emphasis is on "close-to-linear" models, whose parameter estimators are close to being unbiased, normally distributed, minimum variance estimators. Note price. RSK

Statistics, S(18), P. *Bayesian Forecasting and Dynamic Models.* Mike West, Jeff Harrison. Ser. in Stat. Springer-Verlag, 1989, xxi + 704 pp, $56. [ISBN: 0-387-97025-8] Provides a broad introduction to the modelling ideas and practices of Bayesian forecasting in dynamic model environments. Begins with the simplest and most widely used model—the first-order polynomial model. Goes on to cover regression models, dynamic linear models, and time series models, emphasizing sequential learning and forecasting in a Bayesian fashion. Also covers more advanced, recent developments in the field. Covers theory behind Bayesian forecasting and gives several applications. Extensive bibliography. Comments on computer programs FAB and BATS. JJ

Statistics, T(18), S. *Approximate Distributions of Order Statistics With Applications to Nonparametric Statistics.* R.-D. Reiss. Springer-Verlag, 1989, xii + 355 pp, $59. [ISBN: 0-387-96851-2] Mathematically rigorous treatment of the asymptotic dis-

tribution theory of order statistics. Emphasizes accuracy of limit theorems and higher order approximations. Results formulated in terms of variational and Hellinger distance. Explores parametric approximations of nonparametric models of joint distributions of order statistics, as well as nonparametric curve estimation, the bootstrap method, and graphical methods. Solid knowledge of probability and statistics required. Knowledge of measure-theoretic probability necessary for much of book. Very theoretical. JJ

Statistics, T(18: 1), S, P. *Analysis of Binary Data, Second Edition.* D.R. Cox, E.J. Snell. Mono. on Stat. & Appl. Prob., V. 32. Chapman & Hall, 1989, xi + 236 pp, $39.95. [ISBN: 0-412-30620-4] Deals with the analysis of binary data (zero-one data in which an observation takes one of two possible forms—success or failure). Many methods used to model how the probability of success depends on explanatory variables are unified by considering models in which the logistic transform of the probability of success is a linear combination of unknown parameters. This edition contains more details on regression diagnostics and maximum likelihood fitting than the 1969 *First Edition* by D.R. Cox (TR, June-July 1973). Good mix of theory, application, and examples with real data. Assumes knowledge of statistics. Bibliographic notes in each chapter. JJ

Programming, T(13-14: 1, 2). *Pascal: Problem Solving and Program Design, Third Edition.* Elliot B. Koffman. Addison-Wesley, 1989, xix + 837 pp, (P). [ISBN: 0-201-11834-3] Using smaller chapters, this new edition (*First Edition,* TR, June-July 1981; *Second Edition,* TR, January 1986) increases emphasis on its earlier software engineering topics and adds new ones; uses data abstraction concepts throughout, presents additional theoretical aspects of computer science; adds exercises, examples, case studies, programming projects. Suitable for CS1 and much of CS2 and for the programming course prerequisite for ACM's newly-proposed introductory CS sequence. DFA

Programming, T(15-16: 1), S, P. *Programming in ADA, Third Edition.* J.G.P. Barnes. Intern. Comput. Sci. Ser. Addison-Wesley, 1989, xvi + 494 pp, (P). [ISBN: 0-201-17566-5] An introduction to the entire language, by a member of the original Ada design team. Revised from the earlier editions (*First Edition,* TR, February 1983) to conform to the current interpretation of the language. Includes an enlarged language overview, some rearrangements and rewriting, some new sections, many additional exercises (with solutions), and examples. Tutorial style. Very readable. DFA

Programming, S, P, L*. *Smalltalk-80: The Language.* Adele Goldberg, David Robson. Addison-Wesley, 1989, xvii + 585 pp, (P). [ISBN: 0-201-13688-0] A revised and updated version of *Smalltalk-80: The Language and Its Implementation* (TR, January 1985). Most of the material on implementation of the language has been eliminated. AO

Programming, S, P, L. *Real World PostScript: Techniques from PostScript Professionals.* Ed: Stephen F. Roth. Addison-Wesley, 1988, xiv + 383 pp, (P). [ISBN: 0-201-06663-7] A "forager's book" of real opportunities of PostScript as a programming language, including applications to development of type fonts, half-toning and color, and four case studies in major projects. Each chapter is individually authored; most include extensive chunks of working PostScript code as illustrations. Neither comprehensive nor systematic, but full of practical wisdom not available in standard reference manuals. LAS

Programming, S. *HP-28S Software Power Tools: Utilities.* Solve and Integrate Corp. Grapevine Publ (460 SW Madison Ave., Suite 1, POB 118, Corvallis, OR 97339), 1989, 313 pp, $18 (P). [ISBN: 0-931011-27-2] A collection of 200 utility programs for the HP-28S symbolic and graphing calculator. Mathematical examples: finding *gcd* and *lcm*; generating primes and random numbers; converting formats of complex numbers and vectors; building, decomposing, and editing arrays. Other utilities pertain to strings, lists, directories, output, and programming. LAS

Programming, S, P, L. *An Introduction to Object-Oriented Programming and Smalltalk.* Lewis J. Pinson, Richard S. Wiener. Addison-Wesley, 1988, xxvi + 502 pp, (P). [ISBN: 0-201-19127] An introduction to the principles of object-oriented programming, the Smalltalk programming environment, and the Smalltalk-80 programming language. Does not assume that the reader has any prior experience with Smalltalk or object-oriented programming. AO

Programming, S, P, L. *C++ Primer.* Stanley B. Lippman. Addison-Wesley, 1989, xiv + 464 pp. [ISBN: 0-201-16487-6] A tutorial introduction to the programming language C++. Does not assume prior experience with C or C++. Describes Release 2.0 of the language and the differences between Release 1.2 and Release 2.0. AO

Programming, T(13-14: 1), P. *Fortran 8x Explained, Revised Edition.* Michael Metcalf, John Reid. Clarendon Pr, 1989, xv + 315 pp, $22.50 (P). [ISBN: 0-19-853824-3] A presentation of the second draft, not yet adopted, of *Fortran 8x,* the proposed new standard containing Fortran 77 as a subset. Written in textbook style, as if Fortran 8x were in existence. Notes new features and proposed "obsolescent" features. FR

Languages, P, L. *Constraint Programming Languages: Their Specification and Generation.* Wm Leler. Addison-Wesley, 1988, vi + 202 pp, $10.95 (P). [ISBN: 0-201-06243-7] Constraint languages are a relatively new programming paradigm. They are declarative in that they specify a desired goal rather than an algorithm to achieve that goal. This book describes a general-purpose specification language called Bertrand that can be used to produce constraint-satisfaction systems based on a new rewrite rule technique entitled augmented term rewriting. AO

Languages, T(15-17: 1, 2), P, L. *Elements of*

Functional Programming. Chris Reade. Intern. Comput. Sci. Ser. Addison-Wesley, 1989, xv + 600 pp. [ISBN: 0-201-12915-9] A comprehensive introduction to the concepts and techniques used in functional programming. Gives particular emphasis to the use of abstraction and strong polymorphic type systems. Standard ML is used in the examples. AO

Algorithms, T(15-16: 1), L. *Computer Algorithms: Introduction to Design and Analysis, Second Edition.* Sara Baase. Addison-Wesley, 1988, xv + 416 pp. [ISBN: 0-201-06035-3] *Second Edition* of this well-known textbook on algorithm analysis (*First Edition,* TR, February 1979). Additions include a large number of exercises as well as material on adversary arguments and selection problems, dynamic programming, and parallel algorithms. AO

Computer Systems, S(15-17), P, L*. *The Design and Implementation of the 4.3BSD UNIX Operating System.* Samuel J. Leffler, *et al.* Addison-Wesley, 1989, xxii + 471 pp. [ISBN: 0-201-06196-1] Describes the internal structure (concepts, data structures, algorithms, etc.) of the widely used 4.3BSD version of UNIX developed at the University of California at Berkeley. Some exercises are provided to facilitate the use of this book as a supplement in a second course on operating systems design. AO

Computer Systems, P. *Digital Typography: An Introduction to Type and Composition for Computer System Design.* Richard Rubinstein. Addison-Wesley, 1988, xi + 340 pp. [ISBN: 0-201-17633-5] A broad introduction to the technical and aesthetic issues that arise in attempts to produce typeset-quality output with computer-driven equipment. The topics covered include a brief history of letterforms, the psychology of reading, characteristics of various output devices, principles of font design and page layout, and the relationship between screen and paper images. AO

Computer Science, T(18), P. *Algebraic Specification.* Eds: J.A. Bergstra, J. Heering, P. Klint. Addison-Wesley, 1989, xiii + 397 pp. [ISBN: 0-201-41635-2] There is a good deal of research going on in the area of the formal specification of the syntax and semantics of programming languages. This monograph describes the research activities in one area of this research—algebraic specifications. The authors describe a system called ASF which uses algebraic techniques to describe language. They then apply this technique to a number of sample languages to show how it can be used. GMS

Applications, T(13-14: 1). *How to Model It: Problem Solving for the Computer Age.* Anthony M. Starfield, Karl A. Smith, Andrew L. Bleloch. McGraw-Hill, 1990, xii + 206 pp, $19.95 (P). [ISBN: 0-07-005897-0] An informal non-mathematical approach to modelling designed to stimulate active, thoughtful discussion-centered teamwork among students who may not be very comfortable with formal mathematics. Encourages use of spreadsheets as a standard tool. Sample modelling projects: estimating how long it takes to cool a warm beer in a refrigerator; planning to cook a dinner in a limited amount of time; determining how many ping pong balls would fill a room. LAS

Applications (Physics), P. *Random Media and Composites.* Eds: Robert V. Kohn, Graeme W. Milton. SIAM, 1989, 214 pp, $27.50 (P). [ISBN: 0-89871-246-7] Compiled from papers delivered at the workshop on random media and composites held at the Xerox Training Center in Leesburg, Virginia in December 1988. Topics of principle concern include percolation and breakdown in random resistor networks, bounds and mean fields theories for the effective moduli of composites, seismic inverse problems, localization in random media, and superconductivity. MU

Applications (Physics), S(18), P. *Lecture Notes in Physics-344: PDEs and Continuum Models of Phase Transitions.* Eds: M. Rascle, D. Serre, M. Slemrod. Springer-Verlag, 1989, vii + 229 pp, $29.70. [ISBN: 0-387-51617-4] Proceedings of an NSF-CNRS Joint Seminar held in Nice, France, January 1988. The first section, Continuum Mechanics, includes papers on liquid crystals, elastic crystals, fluids, and general theory. The second section, Mathematical Analysis, concerns systems of conservation laws, and miscellaneous topics. MU

Applications (Physics), P. *Boundary Stabilization of Thin Plates.* John E. Lagnese. Stud. in Appl. Math., V. 10. SIAM, 1989, viii + 176 pp, $36.50. [ISBN: 0-89871-237-8] A comprehensive treatment of asymptotic stability of a thin plate when stabilizing feedback mechanisms are introduced along a part of the edge. Primary emphasis is placed on the use of multipliers to obtain explicit estimates of the decay rate of the energy of the plate. LCL

Applications (Social Science), P, L. *Nystuen-Dacey Nodal Analysis.* Ed: Keith J. Tinkler. Mono. No. 7. Institute of Mathematical Geography (2790 Briarcliff, Ann Arbor, MI 48105), 1988, iii + 115 pp, $15.95 (P). A monograph giving detailed instructions for the implementation of Tinkler's method of nodal analysis. Plenty of examples in the appendices. Appendices also include reprints of earlier work by Tinkler, as well as work by John D. Nystuen and Michael F. Dacey, the founders of nodal analysis in geography. SM

Reviewers

DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; WE: William Etter, Macalester; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RWJ: Roger W. Johnson, Carleton; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; AO: Arnold Ostebee, St. Olaf; FR: Flauren Ricketts, Macalester; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; KES: Kay E. Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.

A D D I S O N - W E S L E Y

CALCULUS
like you've never seen it before!

Now you and your students can keep in step with today's colorful new math movement through our innovative calculus texts.

Calculus
Ross L. Finney and George B. Thomas Jr., both of *Massachusetts Institute of Technology*

This text emphasizes a new way of learning and teaching calculus. Finney and Thomas integrate calculators and computers to give students an overall problem-solving and graphing approach to learning.

An extensive supplements package includes these computer supplements:

- **AWTest** (Computerized Testing)
- **Calculus Toolkit 2.0** (Apple and IBM)
- **Computer Explorations in Calculus** (Workbook to accompany the **Calculus Toolkit**)*
- **Mathematical Modeling with MathCAD** (Workbook/disk to accompany **The Student Edition of MathCAD**)*
- **The Student Edition of MathCAD®** (IBM)*
- **Master Grapher** and **3D Grapher** for the IBM, Apple II series, and the Macintosh; in 3 1/2" and 5 1/4" disk sizes.*
- **Mathpro** (IBM)*

Calculus with Analytic Geometry, Third Edition
John B. Fraleigh, *University of Rhode Island*

Fraleigh illustrates topics through graphing and extensive calculator use to give students a clear understanding of calculus. Exponential and logarithmic functions are introduced early, which means your students can apply these concepts throughout the semester. Plus, this edition includes:

- A comprehensive dictionary of functions in the precalculus review chapter.
- **Exploring Calculus,** an IBM tutorial software package.
- **AWTest Edit,** a computerized testing system with editing capabilites.

Calculus and Analytic Geometry, Seventh Edition
George B. Thomas Jr., and Ross L. Finney, both of *Massachusetts Institute of Technology*

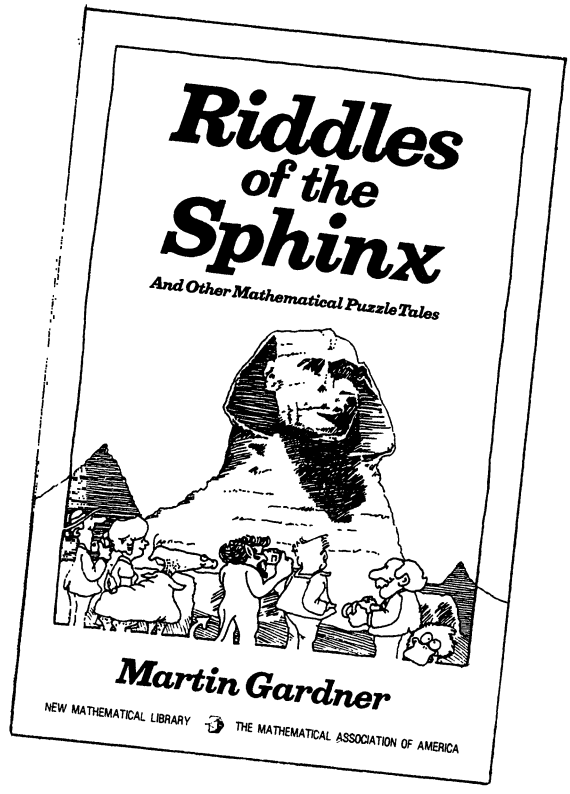
Seventh edition of the classic, best-selling text, known for its excellent applications and mathematical precision.

* These items accompany both **Calculus** and **Calculus with Analytic Geometry, Third Edition.**

Addison-Wesley Publishing Company
1 Jacob Way • Reading, MA 01867
617-944-3700

Riddles of the Sphinx

and other mathematical puzzle tales.



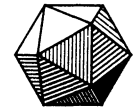
by Martin Gardner
Volume 32 in the New Mathematical Library
184 pp., Paper. ISBN-0-88285-632-8
List: $14.50 MAA Member: $12.50

Martin Gardner has charmed readers for over fifty years with his delightful books and articles on science. He is best known for the popular column, *"Mathematical Games,"* which appeared in *Scientific American* for twenty-five years. Generations of scientists and mathematicians have been inspired by his writing and the MAA is proud to include his name in its list of authors.

This book was drawn from Gardner's column in *Isaac Asimov's Science Fiction Magazine.* The riddles presented here incorporate the responses of his initial readers, along with additions suggested by the editors of the New Mathematical Library. Each chapter (riddle) poses a problem answered in the First Answers section. The solution in turn raises another problem that is solved in the Second Answers section. This may suggest a third question and in several instances there is a fourth. Gardner draws us from questions to answers always presenting us with new riddles—some as yet unanswered. There are 125 different pieces altogether.

Solving these riddles is not simply a matter of logic and calculation, although these play a role. Luck and inspiration are factors as well, so beginners and experts alike may profitably exercise their wits on Gardner's problems, whose subjects range from geometry to word play to questions relating to physics and geology.

We guarantee that you will solve some of the riddles, be stumped by others, and be amused by almost all of the stories and settings that Gardner has devised to raise these questions.



Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036



Fields & Operators

Surfaces. Vector Fields. Differential Operators. In two or three dimensions.

And it's animated!

From our mail...

"Another Success from Lascaux Graphics!" M.E. City College of N Y

"Excellent Program!! ..An exceptional educational aid." D.W. - Dow Chemical

"Just what I've been looking for!" D.B. Olivet College

"Keep up the valuable programs at decent prices." D.W.

"Amazing Software - write more." Dr. R. B. TNU

"Without your program's powerful layering and animation capabilities, we would not have attempted such a task." M. T. Univ. of Wisconsin

"A super product" Recreational & Educational Computing

For PC or Macintosh

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

Writing Mathematics Well, by Leonard Gillman
64 pp., 1987, ISBN-0-88385-443-0
Catalog Number - WMW List: $6.50 MAA Member: $5.00

Good writing conveys more than the author originally had in mind, while poor writing conveys less. Well-written papers are more quickly accepted and put into print and more widely read and appreciated than poorly written ones—and for notes, monographs, and books the quality of writing is more important than it is for papers.

In **Writing Mathematics Well,** Leonard Gillman tells his readers how to develop a clear and effective style. All aspects of mathematical writing are covered, from general organization and choice of title, to the presentation of results, to fine points on using words and symbols, to revision, and finally, to the mechanics of putting your manuscript into print. No book can by itself make you a better writer, but this one will alert you to the opportunities for better and more forceful writing. It does this both by precept and by example.

A book to be read for its sharpness and wit as well as for enlightenment, **Writing Mathematics Well** should be on the shelf of anyone who writes or intends to write mathematics. It will amuse and delight the already careful writer and it will help reform and refine the sensibilities of those who may be somewhat careless about their writing.



Order from
The Mathematical Association of America
1529 Eighteenth St. NW
Washington, D.C. 20036

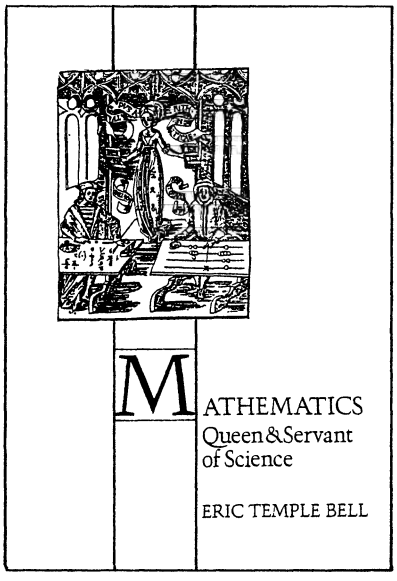
MATHEMATICS: Queen and Servant of Science

by E. T. Bell

"Bell is a lively, stimulating writer, inoffensively crotchety and opinionated, with a good sense of historical circumstance, a fine impatience with humbug, a sound grasp of the entire mathematical scene, and a gift for clear and orderly explanation."
James R. Newman, *The World of Mathematics*

"Bell's writing style makes you feel as if you are a fortunate student spending time with a charming and gifted teacher. Not a history of mathematics, but a wide ranging exposition of ideas in the subject, the book is a classic and should be required reading for every teacher of mathematics."
Harold Jacobs, Grant High School, Van Nuys, California

"This book deserves a place in today's market. It is a much more popular work than most histories of the subject, and that is exactly what makes it accessible to students as well as to non-mathematicians. It is rewarding reading for . . . teachers and students of mathematics at all levels."
Morris Kline, Courant Institute



xv + 437 pp. Paperbound
Republished 1987.
ISBN-0-88385-446-3
List: $18.50
MAA Member: $14.00

This is an absorbing account of the development in pure and applied mathematics from the geometry of Euclid more than two-thousand years ago to some of the more recent developments in mathematical physics. Published 30 years ago, MATHEMATICS: QUEEN AND SERVANT OF SCIENCE remains a valuable tool for teachers, students, and all those enthusiastic about mathematics. The book includes 20 chapters on selected areas in pure and applied mathematics, including discussions on topics such as geometry, algebra, logic, probability, infinity, transformations, rings, groups, topology, and continuity. Also covered are subjects such as: Mersenne primes, quantum mechanics, and three-valued logic. The book is not a history of mathematics, but rather an enticing sampling of selected topics in mathematics and science.

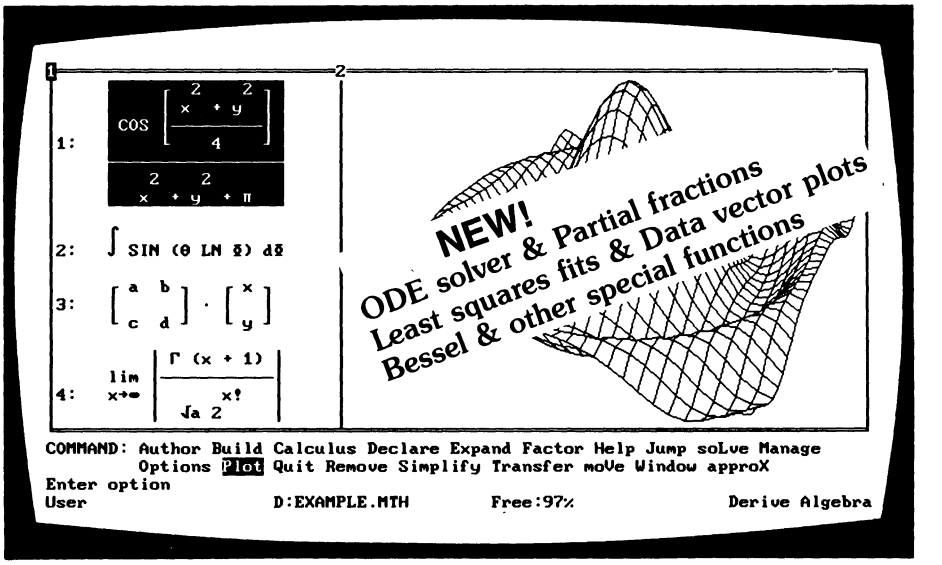
The book contains fascinating information on the personalities and philosophies of the great mathematicians, plus perceptive resumes of their specific contributions to the moving force of this great science.



Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

DERIVE

A Mathematical Assistant



2000 years of mathematical knowledge on a disk

DERIVE, the successor to **muMATH**, is a powerful computer algebra system for your PC compatible computer that provides the following capabilities:

- Exact and approximate arithmetic to thousands of digits
- Equations, complex numbers, trigonometry, calculus, vectors, and matrices
- 2D and 3D function plotting with zooming capability
- MDA, CGA, EGA, VGA, and Hercules graphics and text support
- Attractive 2D mathematical display of formulas
- Easy to use menu-driven interface with on-line help
- Ideal for engineers, scientists, students and teachers
- $200 plus shipping: **Call or write** for information.

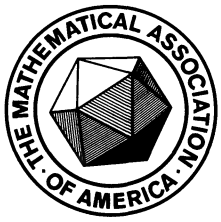
System requirements: IBM PC or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (360K) or a 3½ inch (760K) diskette drive. **Or** NEC PC-9801 or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (640K) diskette drive.

DERIVE and muMATH are trademarks of Soft Warehouse, Inc. Hercules is a trademark of Hercules Computer Technology, Inc. IBM is a registered trademark of International Business Machines Corp. MS-DOS is a registered trademark of Microsoft Corp. NEC is a registered trademark of Nippon Electric Company.

Soft Warehouse INC
3615 Harding Avenue, Suite 505 • Honolulu, HI 96816
(808) 734-5801 after noon PST

Handcrafted software for the mind.

©1988 Soft Warehouse, Inc.



Publications by the
Mathematical Association of America
1529 Eighteenth St., NW
Washington, DC 20036
202 387-5200

New Mathematical Library #30 . . .

The Role of Mathematics in Science
by Leon Bowden and M.M. Schiffer
220 pp. Paperbound
List: $16.00 MAA Member: $13.00

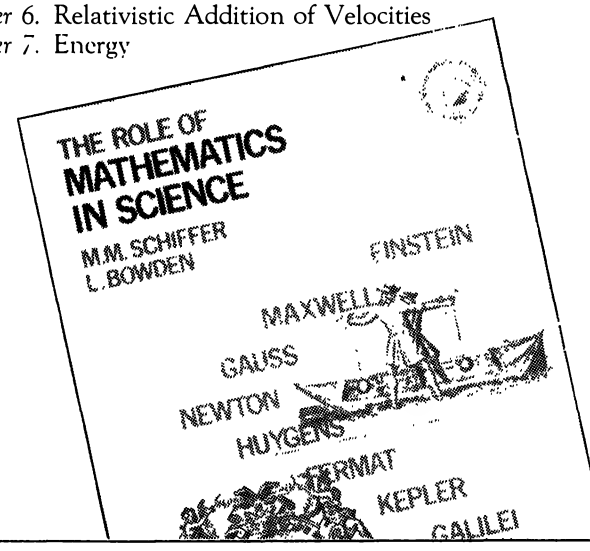
This book is based on a series of lectures given over a period of years to high school mathematics teachers. The lectures focused on topics in mechanics, questions of population growth, probability and other uses of exponential functions, optics and application of matrices to relativity theory.

The aim of the book is to illustrate the power and elegance of mathematical reasoning in science with some examples ranging from the work of Archimedes to that of Einstein. The book starts with problems of the lever, the mirror and the growth of populations and ends up with problems of space travel and atomic energy.

This excellent book will appeal to the advanced high school student, to the undergraduate and to their teachers. A sampling of the Table of Contents should inspire you to order your copy now.

Table of Contents

Chapter 1. The Beginnings of Mechanics
Chapter 2. Growth Functions
Chapter 3. The Role of Mathematics in Optics
Chapter 4. Mathematics with Matrices - Transformations
Chapter 5. What is the Time? Einstein's Transformation Problem
Chapter 6. Relativistic Addition of Velocities
Chapter 7. Energy



New. . .

Writing Mathematics Well, by Leonard Gillman
64 pp., 1987, ISBN-0-88385-443-0
Catalog Number - WMW List: $6.00 MAA Member: $4.50

Good writing conveys more than the author originally had in mind, while poor writing conveys less. Well-written papers are more quickly accepted and put into print and more widely read and appreciated than poorly written ones—and for notes, monographs, and books the quality of writing is more important than it is for papers.

In **Writing Mathematics Well,** Leonard Gillman tells his readers how to develop a clear and effective style. All aspects of mathematical writing are covered, from general organization and choice of title, to the presentation of results, to fine points on using words and symbols, to revision, and finally, to the mechanics of putting your manuscript into print. No book can by itself make you a better writer, but this one will alert you to the opportunities for better and more forceful writing. It does this both by precept and by example.

A book to be read for its sharpness and wit as well as for enlightenment, **Writing Mathematics Well** should be on the shelf of anyone who writes or intends to write mathematics. It will amuse and delight the already careful writer and it will help reform and refine the sensibilities of those who may be somewhat careless about their writing.



Order from
The Mathematical Association of America
1529 Eighteenth St. NW
Washington, D.C. 20036

BE DISCRETE

generatingfunctionology

Herbert S. Wilf

Generating functions are a "bridge" between discrete mathematics on the one hand, and continuous analysis and complex variable theory on the other. This book is about generating functions and some of their uses in discrete mathematics. It is designed for use at the advanced undergraduate level for courses in discrete mathematics and it gives a good introduction to most of the basic enumerative applications of generating functions in a very readable form. Each chapter includes exercises with solutions.

CONTENTS: Introductory Ideas and Examples. Series. Cards, Decks, and Hands. Applications of Generating Functions. Analytic and Asymptotic Methods. Each chapter includes exercises. Solutions.

January 1990, 196 pages, $29.95
ISBN: 0-12-751955-6

Order from your local bookseller or directly from



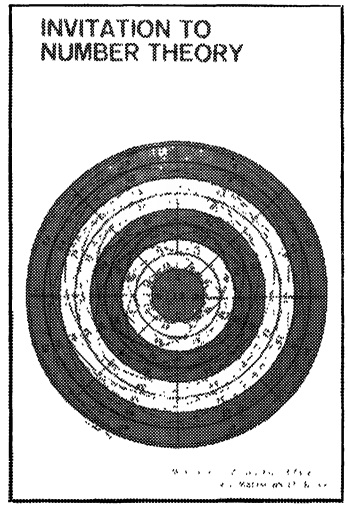
ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
Book Marketing Dept. #12040
1250 Sixth Ave., San Diego, CA 92101

CALL TOLL FREE
1-800-321-5068

Quote this reference number for *free postage and handling* on your prepaid order ➥ **12040**

Prices subject to change without notice. ©1990 by Academic Press, Inc
All Rights Reserved. TC/MJD —12040

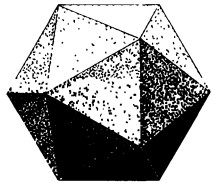
FROM THE NEW MATHEMATICAL LIBRARY



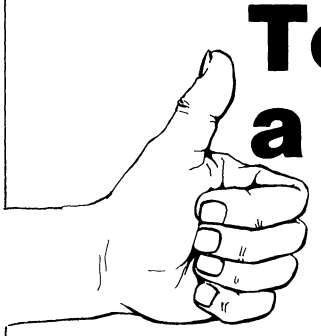
Invitation to Number Theory,
by Oystein Ore
129 pp., 1967, Paper, ISBN-0-88385-620-4
List: $9.90 MAA Member: $7.90

This outstanding book gives the reader some of the history of number theory, touching on triangular and pentagonal numbers, magic squares and Pythagorean triples, and numeration systems. It covers the primes and prime factorization (including the fundamental theorem of arithmetic), congruences (modular arithmetic) and their applications (including methods of checking numerical calculations), tests for primality, scheduling tournaments, and ways of determining the week day of a given date.

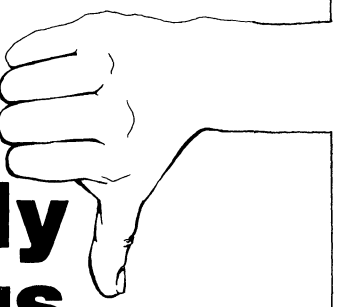
Ore writes of his book, "The purpose of this simple little guide will have been achieved if it should lead some of its readers to appreciate why the properties of numbers can be so fascinating. It would be better still if it would induce you to try to find some number relations of your own; new curiosities devised by young people turn up every year." The enterprise of making such discoveries is very broad including the invention and study of the curious sequence, 1, 11, 21, 1211, 111221, . . . (to understand it read it aloud) by Cambridge Professor John Horton Conway to the discovery of new Mersenne primes by high school students Laura Nickel and Kurt Noll. Ore's book is a good place for readers to learn of the fascination that numbers hold.



Order from:
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036



Toward a Lean and Lively Calculus

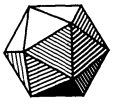


Report of the Conference/Workshop to Develop Curriculum and Teaching Methods for Calculus at the College Level, Ronald G. Douglas, Editor.
MAA Notes #6

249 pp., 1987, Paperbound, ISBN-0-88385-056-7
Catalog Number - NTE-07
Price: $12.50

Should calculus be taught differently? Can it? Common wisdom says "no"—which topics are taught, and when, are dictated by the logic of the subject and by client department. The surprising answer from a four-day Sloan Foundation-sponsored conference on calculus instruction, chaired by Ronald Douglas, is that significant change is possible, desirable, and necessary. Meeting at Tulane University in New Orleans in January, 1986, a diverse and sometimes contentious group of twenty-five faculty, university and foundation administrators, and scientists from client departments, put aside their differences to call for a leaner, livelier, more contemporary course, more sharply focused on calculus's central ideas and on its role as the language of science.

This volume contains the results of that conference and the papers presented to the conferees. These are certain to be the point of departure and basis for efforts to strengthen and reshape calculus in the next decade.



Order from
The Mathematical Association
of America
1529 Eighteenth St., NW
Washington, DC 20036

MAA STUDIES IN MATHEMATICS

Studies in Numerical Analysis

MAA Studies in Mathematics #24

Gene H. Golub, Editor

415 pp. Hardbound.

List: $46.50 MAA Member: $34.50

This volume is a collection of papers describing the wide range of research activity in numerical analysis. The articles describe solutions to a variety of problems using many different kinds of computational tools. Some of the computations require nothing more than a hand held calculator: others require the most modern computer. While the papers do not cover all of the problems that arise in numerical analysis, they do offer an enticing and informative sample.

Numerical analysis has a long tradition within mathematics and science, beginning with the work of the early astronomers who needed numerical procedures to help them solve complex problems. The subject has grown and developed many branches, but it has not become compartmentalized. Solving problems using numerical techniques often requires an understanding of several of the branches. This fact is reflected in the papers in this collection.

Computational devices have expanded tremendously over the years, and the papers in this volume present the different techniques needed for and made possible by several of these computational devices.

Table of Contents

The Perfidious Polynomial, *James H. Wilkinson*

Newton's Method, *Jorge J. Moré and D. C. Sorensen*

Research Directions in Sparse Matrix Computations, *Iain S. Duff*

Questions of Numerical Conditions Related to Polynomials, *Walter Gautschi*

A Generalized Conjugate Gradient Method for the Numerical Solution of Elliptic Partial Differential Equations, *Paul Concus, Gene H. Golub and Dianne P. O'Leary*

Solving Differential Equations on a Hand Held Programmable Calculator. *J. Barkley Rosser*

Finite Difference Solution of Boundary Value Problems in Ordinary Differential Equations, *V. Pereyra*

Multigrid Methods for Partial Differential Equations, *Dennis C. Jespersen*

Fast Poisson Solvers, *Paul N. Swarztrauber*

Poisson's Equation in a Hypercube: Discrete Fourier Methods, Eigenfunction Expansions, Pade Approximation to Eigenvalues, *Peter Henrici*



Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 5 May 1990

Contents

(ISSN 0002-9890)

ARTICLES

The Ergodic Character of Sequences of Pedal Triangles PETER D. LAX 377

Stability Properties of Geometric Inequalities H. GROEMER 382

NOTES

Separable Extensions and Diagonalizability FRED RICHMAN 395

A Birthday Present for Ramanujan M. D. HIRSCHHORN 398

A Condition for Flatness of Curves in R^n YAAKOV S. KUPITZ AND MICHA A. PERLES 401

A Counterexample on Vector Norms and the Subordinate Matrix Norms PIETER P. N. DE GROEN 406

Ambiguous Numbers are Dense MARKO PETKOVŠEK 408

The Diophantine Equation $x^2 + y^m = z^{2n}$, Addendum DAVID W. BOYD 411

THE TEACHING OF MATHEMATICS

Singular Values and the Spectral Theorem K. HOECHSMANN 413

A Remark on the Shortest Confidence Interval of a Normal Mean ROBERT BARTOSZYŃSKI AND WAI CHAN 415

For Your Eyes Only GARY A. MARTIN 418

Cauchy-Bunyakowski-Schwarz Inequality Revisited FRANÇOIS DUBEAU 419

An Old Max-Min Problem Revisited MARY EMBRY-WARDROP 421

Functional Equations and L'Hôpital's Rule in an Exact Poisson Derivation J. ACZÉL 423

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 427

Advanced Problems and Solutions 432

REVIEWS

The Mathematics of Plato's Academy. A New Reconstruction. By D. H. Fowler RICHARD K. GUY 440

TELEGRAPHIC REVIEWS 444

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

The Ergodic Character of Sequences of Pedal Triangles

PETER D. LAX

PETER D. LAX received his Ph.D. from the Courant Institute of Mathematical Sciences of New York University in 1949 and has been on its faculty since 1950, doing research in both pure and applied mathematics. The present article is an application of concepts from applied mathematics to a problem in pure mathematics.



In an interesting article in this journal, Kingston and Synge discuss sequences of pedal triangles $\{T_n\}$, where T_{n+1} is the pedal triangle of T_n, i.e., its vertices are the feet of the altitudes of T_n. The authors observe that there are pedal sequences which are periodic in shape, and construct all of them. In this note we demonstrate that the shape of triangles in a pedal sequence is ergodic.

Denote by A, B, C the angles of a given triangle; the angles A', B', C' of its pedal triangle are simply related to A, B, C:

If all angles A, B, C are acute,

$$A' = \pi - 2A, \qquad B' = \pi - 2B, \qquad C' = \pi - 2C; \tag{1}$$

if A is obtuse,

$$A' = 2A - \pi, \qquad B' = 2B, \qquad C' = 2C, \tag{2}$$

with analogous formulas when B and C are obtuse. Let's call the mapping

$$A, B, C \to A', B', C'$$

the *pedal mapping*, denoted by **P**.

Kingston and Synge give an elegant and useful representation to the pedal mapping renormalized by dividing all angles by π, i.e., setting

$$A = a\pi \text{ etc.}, \qquad A' = a'\pi \text{ etc.}$$

For then

$$a + b + c = 1, \qquad a' + b' + c' = 1,$$

so that (a, b, c) and (a', b', c') can be regarded as barycentric coordinates of points of a triangle E, with vertices P, Q, R. The pedal mapping can then be regarded as a mapping of E into E. Divide E into four congruent triangles, each half the size of E, see FIGURE 1. Denote the subtriangles adjacent to P, Q, R by E_i, $i = P, Q, R$, and denote by E_M the one in the middle.

A triangle T whose angles are acute corresponds to a point interior to E_M. Formulas (1) and (2) for the angles of the pedal triangle can be expressed as follows:

P maps points of E_M by dilatation from the midpoint $(1/3, 1/3, 1/3)$ of E by a scale factor -2.

P maps points of E_i, $i = P, Q, R$ by dilatation from P, Q or R by a scale factor 2.

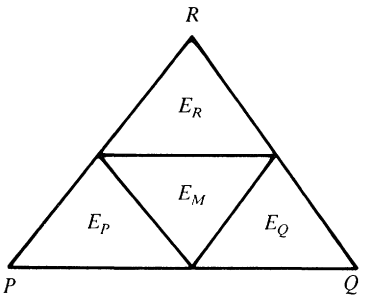


FIG. 1

Note that each E_i is mapped onto E; $\mathbf{P}$ is discontinuous across the boundary of E_M, except for the midpoints of its sides.

The pedal mapping can be regarded as a two-dimensional analogue of the mapping $x \to f(x)$ of the unit interval into itself defined by

$$f(x) = \begin{cases} 2x & 0 \leqslant x \leqslant \frac{1}{2} \\ 2 - 2x & \frac{1}{2} \leqslant x \leqslant 1. \end{cases} \tag{3}$$

The graph of f is roof-shaped:

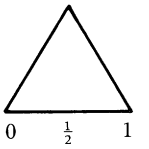


FIG. 2

In fact, the restriction of the pedal map to any of the medians of E is the roof mapping (3).

The pedal mapping shares many of the properties of the roof mapping:

THEOREM 1. *The pedal mapping is area preserving, i.e., for any measurable subset G of E,*

$$m(G) = m(\mathbf{P}^{-1}(G)),$$

where m is Lebesgue measure in the plane.

Proof. This is obvious, since the inverse image of G consists of four pieces G_i located in the four subtriangles E_i, $i = P, Q, R$, and M, each congruent to G diminished by a factor of 2. Therefore each G_i has measure $(1/4)m(G)$, and their union has measure $m(G)$.

THEOREM 2. *The pedal mapping is ergodic.*

This means: for any point q in E, any subset O of E and any natural number n denote by

$$N(q, O, n)$$

the number of points

$$q, \mathbf{P}(q), \ldots, \mathbf{P}^{n-1}(q)$$

that lie in O. The ergodic property is that for almost all q and every open set O,

$$\lim_{n\to\infty} \frac{N(q, O, n)}{n} = \frac{\text{area of } O}{\text{area of } E}.$$

In words: for almost all q, the successive images of q under $\mathbf{P}$ are on the average equally distributed over E.

According to G. D. Birkhoff's pointwise ergodic theorem, a measure preserving map is ergodic iff it is metrically transitive. This means any measurable subset of E that is invariant under $\mathbf{P}$ has measure either 0 or $m(E)$.

We show now that the pedal map $\mathbf{P}$ is metrically transitive. Let $S \subset E$ be a set invariant under $\mathbf{P}$; denote by c its relative measure

$$c = \frac{m(S)}{m(E)}. \tag{4}$$

Denote by S_i the portion of S contained in E_i, $i = P, Q, R, M$, *i.e.*, $S_i = S \cap E_i$.

Since S is invariant under $\mathbf{P}$, the images $\mathbf{P}S_i$ are contained in S. Since the measure of $\mathbf{P}G$ is four times the measure of G for any set G contained in any of the subtriangles E_i, it follows that

$$4m(S_i) \leqslant m(S), \qquad i = P, Q, R, M. \tag{5}$$

Adding these inequalities and dividing by 4 we get

$$\sum m(S_i) \leqslant m(S).$$

Using the additivity of measure, and that

$$\cup S_i = S,$$

we get

$$m(S) \leqslant m(S).$$

It follows that the sign of equality must hold in (5) for all i:

$$m(S_i) = \frac{1}{4} m(S).$$

Combining this with the notation (4), and using the fact that

$$m(E_i) = \frac{1}{4} m(E),$$

we get that

$$m(S_i) = cm(E_i), \tag{6}$$

for $i = P, Q, R, M$.

We claim that (6) holds for every triangle E_i obtained from E by repeated subdivision into four congruent parts. We have already shown this for the four triangles obtained by a single subdivision; to show it for triangles obtained by n subdivisions we argue by induction on n. Denote by $\{E_i'\}$ all the triangles obtained by $n+1$ subdivisions, and denote by S_i' the intersection of E_i' and S. **P** maps each E_i' onto a triangle E_j, obtained by n subdivisions; therefore S_i' is contained in $S_j = E_j \cap S$. Arguing as in (5) we conclude that

$$4m(S_i') \leqslant m(S_j).$$

By induction hypothesis, (6) holds for S_j; therefore,

$$4m(S_i') \leqslant cm(E_j) = c4m(E_i'). \tag{7}$$

We sum over all S_i'; using additivity of measure and that

$$\cup S_i' = S, \qquad \cup E_i' = E$$

we obtain, after dividing by four,

$$m(S) \leqslant cm(E).$$

According to (4) the sign of equality holds above; therefore equality must hold in (7) for all i; this completes the induction.

Next we claim that (6) holds for all rectangles R in E, i.e., that for all R,

$$m(R \cap S) = cm(R). \tag{8}$$

For every rectangle can be approximated arbitrarily closely from the inside and outside by unions of triangles E_i.

Denote by $R(h, p)$ the square of side length h centered at p. According to a classical theorem of Lebesgue, for any measurable set S,

$$\lim_{h \to 0} \frac{m(R(h, p) \cap S)}{m(R(h))}$$

exists for almost all p, and the limit is 1 at almost all points of S, and zero at almost all points in the complement of S. In light of (8) this means that c is either 1 or 0. In the first case

$$m(S) = m(T),$$

in the second case

$$m(S) = 0.$$

This completes the proof that **P** has no nontrivial invariant sets, and thereby has the ergodic property. Q.E.D.

We can simplify the last step in the proof by appealing to the much more elementary L^1 version of Lebesgue's theorem, as follows. Denote by ρ the characteristic function of the unit square R around the origin, i.e.,

$$\rho(p) = \begin{cases} 1 & \text{for } p \text{ in } R \\ 0 & \text{otherwise.} \end{cases}$$

Denote by ρ_h the function

$$\rho_h(p) = h^{-2}\rho(ph^{-1});$$

for small h, ρ_h serves as an approximate δ-function. For any χ in L^1 we denote by χ_h its mollification by ρ_h

$$\chi_h = \rho_h * \chi,$$

where $*$ denotes convolution. Obviously, if χ is continuous, χ_h tends uniformly to χ as h tends to zero. Since convolution with ρ_h is bounded in the L^1 norm

$$|\rho_h * \chi|_1 \leq |\chi|_1,$$

it follows that for any χ in L^1, χ_h tends to χ in the L^1 norm as h tends to zero.

We apply this to the characteristic function χ of S

$$\chi(p) = \begin{cases} 1 & \text{for } p \text{ in } S \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

It follows from (8) that for any h,

$$\chi_h \equiv c.$$

Taking the limit as h tends to zero we conclude that $\chi \equiv c$. In view of (9) we conclude that c is either 0 or 1, and S has measure either 0 or 1. Q.E.D.

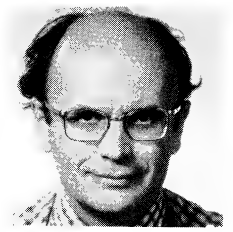
REFERENCES

1. P. R. Halmos, Lectures on Ergodic Theory, The Mathematical Society of Japan, 1956.
2. J. G. Kingston and J. L. Synge, The sequence of pedal triangles, *Amer. Math. Monthly*, 95 (1988) 609–622.

Stability Properties of Geometric Inequalities

H. GROEMER,* *University of Arizona, Tucson, AZ 85721*

HELMUT GROEMER studied mathematics at the Universities of Innsbruck and Vienna. He received his Ph.D. in 1954 as a student of Leopold Vietoris. In 1957 he came to the United States to join the faculty at Oregon State University. Since 1964 he has been at the University of Arizona. His field of special interest is the theory of convex sets, including packing and covering problems, integral geometry, and stability questions.



Abstract. Geometric inequalities, like the isoperimetric inequality, have been a subject of intensive research for a long time. More recently, some properties of these inequalities, which may be called stability properties, have been investigated. Roughly speaking, these investigations concern the geometric implications if the inequalities are in a certain sense close to equalities. The present article is primarily an exposition of various stability results concerning inequalities for plane convex sets, including a discussion of the concepts and methods which are of importance in this area. Moreover, several new results and proofs are presented.

1. Introduction. Let R^n denote n-dimensional Euclidean space. A bounded convex subset of R^n will be called an n-dimensional *convex body* if it is closed and has interior points. We let $\mathscr{C}^n$ denote the class of all n-dimensional convex bodies. Two-dimensional convex bodies will be called *convex domains*.

We consider primarily inequalities of the form

$$\Phi(K) \geq 0, \tag{1}$$

where Φ is real valued and (1) is supposed to hold for all $K \in \mathscr{C}^2$. For example, if we let $p(K)$ denote the perimeter and $a(K)$ the area of K we may set $\Phi(K) = p(K)^2 - 4\pi a(K)$. Then (1) is the well-known isoperimetric inequality.

Let now $\mathscr{C}_\Phi$ denote those elements $K \in \mathscr{C}^2$ for which

$$\Phi(K) = 0.$$

For instance, in the case of the isoperimetric inequality $\mathscr{C}_\Phi$ is known to consist of all (closed) circular discs. We are interested in the stability problem associated with (1). That means, we ask if K must be close to a member of $\mathscr{C}_\Phi$ whenever $\Phi(K)$ is close to 0. In the case of the previous example the question is whether K must be close to a circular disc if $p(K)^2 - 4\pi a(K)$ is small. In order to give a precise formulation of the stability problem let us assume that we are given a function that measures in some sense the deviation between two convex domains. This will always be a nonnegative function $g(K, L)$ defined for all $K, L \in \mathscr{C}^2$ which has the property that $g(K, L) = 0$ if and only if $K = L$. If Φ, $\mathscr{C}_\Phi$ and g are given, the stability problem associated with the geometric inequality (1) can now be formulated as follows:

Find positive constants c, α with the property that whenever

$$\Phi(K) \leq \varepsilon \tag{2}$$

*Supported by National Science Foundation Research Grant DMS 8701893.

(*for some* $\varepsilon \geq 0$), *then there exists an* $L \in \mathscr{C}_\Phi$ *such that*

$$g(K, L) \leq c\varepsilon^\alpha. \tag{3}$$

(Instead of $c\varepsilon^\alpha$ one could admit more general functions but for the purpose of this article nothing would be gained.) For example, if again $\Phi(K) = p(K)^2 - 4\pi a(K)$ and if $g(K, L)$ denotes the Hausdorff distance (as defined in the next section) it will be seen that (2) implies that there is a circular disc L with $g(K, L) \leq (1/4\sqrt{\pi})\varepsilon^{1/2}$ (cf. the remarks regarding (13)).

It is of interest to note that a solution of the above stability problem implies the original inequality (1). Indeed, if we had $\Phi(K) < 0$ we could take $\varepsilon = 0$ which, because of (3) would imply $g(K, L) = 0$. But this is impossible since it would yield $K = L$ and therefore $\Phi(K) = 0$.

If one chooses in (2) $\varepsilon = \Phi(K)$, then $g(K, L) \leq c\Phi(K)^\alpha$ and consequently

$$\Phi(K) \geq (g(K, L)/c)^{1/\alpha}. \tag{4}$$

This inequality, which has to be interpreted in the sense that for every K there is a $L \in \mathscr{C}_\Phi$ such that (4) holds, is obviously a strengthened version of (1). It is also clear that any inequality of the form (4) leads immediately to a solution of the above stability problem. In the case of the isoperimetric inequality the inequality corresponding to (4) and the previously mentioned stability result is

$$p(K)^2 - 4\pi a(K) \geq 16\pi g(K, L)^2,$$

where g is again the Hausdorff distance and L is a suitable circular disc.

Obviously, one can consider many variations and generalizations of these stability problems. For example, there is no need to restrict attention to the case $n = 2$ or to inequalities involving only one convex body. Some of these generalizations will be discussed later.

It is the aim of the present article to survey a variety of stability results. At the same time this exposition provides us with the opportunity to introduce and discuss some of the concepts that proved to be useful in this context. Finally, several new results and proofs will be presented. These mostly concern stability results for plane convex sets that are stronger than those obtained by specializing known results for n-dimensional convex bodies to the case $n = 2$.

2. The Support Function. Let K be a convex domain and assume that the origin $\mathbf{o}$ of R^n is an interior point of K. If $\mathbf{u}$ is a unit vector in R^2 we let $L(\mathbf{u})$ denote the line that is perpendicular to $\mathbf{u}$ and has the property that it meets K but not the interior of K, and that $\mathbf{u}$ points into the halfplane determined by $L(\mathbf{u})$ that does not contain K. We call $L(\mathbf{u})$ the *support line* of K associated with $\mathbf{u}$. The distance between $\mathbf{o}$ and $L(\mathbf{u})$, considered as a function of $\mathbf{u}$, is called the *support function* of K. It is usually denoted by $H(\mathbf{u})$. However, since $\mathbf{u}$ is uniquely determined by the angle, say θ, between the positive x-axis and $\mathbf{u}$ we also write $H(\theta)$ instead of $H(\mathbf{u}(\theta))$. It is easy to see that $H(\mathbf{u}) = \sup\{\mathbf{u} \cdot \mathbf{x} : \mathbf{x} \in K\}$, where $\mathbf{u} \cdot \mathbf{x}$ denotes the usual scalar product. This formula can also serve as a definition of $H(\boldsymbol{u})$ in the case when $\mathbf{o} \notin K$. Obvious arguments show that $H(\theta)$ is a continuous function of θ. The support function is one of the most fundamental concepts in the theory of convex sets and is discussed in practically all textbooks on convexity (see for example [**2**], [**4**], [**10**], [**20**], [**25**]).

As before, let $p(K)$ denote the perimeter and $a(K)$ the area of K. We have

$$p(K) = \int_0^{2\pi} H(\theta)\, d\theta \tag{5}$$

and, if H is sufficiently smooth and $'$ denotes the derivative with respect to θ,

$$a(K) = \frac{1}{2}\int_0^{2\pi} \left(H(\theta)^2 - H'(\theta)^2\right) d\theta. \tag{6}$$

(5) is an immediate consequence of the two-dimensional version of the so-called surface area formula of Cauchy (see [**4**, p. 48]). If K is strictly convex and $H(\theta)$ is continuously differentiable the following analytic proof of (5) is also of interest (cf. Minkowski [**27**]). If we set $\mathbf{q} = K \cap L(\mathbf{u})$ then $H(\mathbf{u}) = \mathbf{q} \cdot \mathbf{u}$ and consequently $H''(\theta) = \mathbf{q}' \cdot \mathbf{u} + \mathbf{q} \cdot \mathbf{u}' = \mathbf{q} \cdot \mathbf{u}'$ (since $\mathbf{q}'$ is a tangent vector of the curve ∂K and therefore perpendicular to $\mathbf{u}$). Hence, $\mathbf{q} = (\mathbf{q} \cdot \mathbf{u})\mathbf{u} + (\mathbf{q} \cdot \mathbf{u}')\mathbf{u}' = H(\theta)\mathbf{u} + H'(\theta)\mathbf{u}'$, and because of $\mathbf{u}'' = -\mathbf{u}$ it follows that $\mathbf{q}'(\theta) = (H(\theta) + H''(\theta))\mathbf{u}'$ where $H(\theta) + H'(\theta) \geq 0$. Thus,

$$p(K) = \int_0^{2\pi} \|\mathbf{q}'(\theta)\|\, d\theta = \int_0^{2\pi} (H(\theta) + H''(\theta))\, d\theta = \int_0^{2\pi} H(\theta)\, d\theta.$$

Next we will prove (6). Assume that H is twice continuously differentiable. If we set $\|\mathbf{q}\| = \mathbf{r}$ and let β denote the angle between the positive x-axis and the vector $\mathbf{q}$, then

$$a(K) = \frac{1}{2}\int_0^{2\pi} r^2\, d\beta.$$

Because of $r\cos(\beta - \theta) = H(\theta)$ and $r = \sqrt{H(\theta)^2 + H'(\theta)^2}$ we obtain after a simple calculation $r^2(d\beta/d\theta) = H(\theta)^2 + H(\theta)H''(\theta)$. Therefore,

$$a(K) = \frac{1}{2}\int_0^{2\pi} \left(H(\theta)^2 + H(\theta)H''(\theta)\right) d\theta,$$

and (6) follows by an obvious integration by parts.

For further comments on (6) see [**4**] and Blaschke [**1**] who has shown that (6) holds for any convex domain if H' is properly interpreted. However for the purpose of proving stability results smoothness assumptions are usually no serious restriction since the general case can be obtained by routine approximation arguments.

Another concept that can be defined in terms of the support function and will play a role in our stability statements is the *Steiner point* of K. It can be defined by

$$\mathbf{s}(K) = \frac{1}{\pi}\int_0^{2\pi} \mathbf{u}(\theta)H(\theta)\, d\theta,$$

where an obvious generalization of the ordinary integral to vector-valued functions is used. If the curve $C = \partial K$ is sufficiently smooth and if one assigns to each point of C the curvature as density, then $\mathbf{s}(K)$ is the center of mass of this "inhomogeneous" curve. For more information on the Steiner point of a convex body one may consult [**19**], [**25**], [**30**], and [**31**].

3. Fourier Series. The support function $H(\theta)$ of a given convex domain K is always continuous, bounded, and of period 2π. Consequently it has a Fourier series

of the form

$$a_0 + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta).$$

We wish to express $p(K)$, $\mathbf{s}(K)$, and $a(K)$ in terms of the Fourier coefficients. From (5) and the definition of the Steiner point and the Fourier coefficients we obtain immediately

$$p(K) = 2\pi a_0 \tag{7}$$

and

$$\mathbf{s}(K) = (a_1, b_1). \tag{8}$$

From (6) and Parseval's equation

$$\int_0^{2\pi} H(\theta)^2 \, d\theta = \pi\left(2a_0^2 + \sum_{n=1}^{\infty} (a_n^2 + b_n^2)\right),$$

together with the corresponding equation for $H'(\theta)$ (if this function is also continuous), one can deduce that

$$a(K) = \pi a_0^2 - \frac{\pi}{2} \sum_{n=2}^{\infty} (n^2 - 1)(a_n^2 + b_n^2). \tag{9}$$

Blaschke [**1**] has shown that (9) can actually be proved without any smoothness assumptions; he also noted the validity of (8). Regarding (8) see also Kubota [**24**], Meissner [**26**], and the historical remarks of Schneider in [**31**].

4. Deviation Measures. Let K and L be two convex domains with respective support functions H_K and H_L. The most frequently used function to measure the deviation of L from K (or K from L) is the Hausdorff distance. It can be defined by

$$h(K, L) = \max_u |H_L(u) - H_K(u)|.$$

The dominating role of the Hausdorff distance in the theory of convexity is evident from such books as [**2**], [**4**], [**10**], [**20**], and [**25**]. However various other deviation measures have been proposed and used (cf. [**33**], [**34**], and the survey [**18**]). One such measure which appears to be of particular value with respect to stability problems is the measure that corresponds to the L_2-metric in function spaces. It is defined by

$$h_2(K, L) = \left(\int_0^{2\pi} |H_L(\theta) - H_K(\theta)|^2 \, d\theta\right)^{1/2}$$

From the continuity of the support function one deduces easily that $h_2(K, L) = 0$ if and only if $K = L$. In the next section it will be seen that h_2 arises in a very natural way; see also [**5**] and [**23**]. The quantitative relationship between h and h_2 has been studied by Vitale [**34**]. Among other (and much more general) results he found that for any two convex domains K, L whose diameters do not exceed d the following relationship is always satisfied:

$$\frac{1}{\sqrt{2\pi}} h_2(K, L) \le h(K, L) \le \left(\frac{3d}{2}\right)^{1/3} h_2(K, L)^{2/3}.$$

If K is given we can ask which circular disc yields the best approximation of K in terms of the h_2 deviation measure. The following theorem shows that this problem leads again to the Steiner point of K. For the sake of convenient formulation we associate with each $K \in \mathscr{C}^2$ the circular disc $S(K)$ which has radius $p(K)/2\pi$ and center at the Steiner point $\mathbf{s}(K)$ of K. We call $S(K)$ the *Steiner disc* of K.

THEOREM 1. *If $K \in \mathscr{C}^2$ and X is any circular disc, then*

$$h_2(K, X)^2 = h_2(K, S(K))^2 + h_2(S(K), X)^2, \tag{10}$$

where $S(K)$ is the Steiner disc of K. In particular, $h_2(K, X)$, considered as a function of X is minimal if and only if X is the Steiner disc of K.

Proof. We only have to prove (10) and may assume that $\mathbf{s}(K) = \mathbf{o}$. Then, because of (7) and (8) the support functions H_K, H_X, and $H_{S(K)}$ have, respectively, the following Fourier series:

$$(p(K)/2\pi) + \sum_{n=2}^{\infty} (a_n \cos n\theta + b_n \sin n\theta),$$

$$(p(X)/2\pi) + \alpha_1 \cos\theta + \beta_1 \sin\theta,$$

$$p(K)/2\pi.$$

Hence, using Parseval's equality we can deduce that

$$h_2(K, X)^2 = \frac{1}{2\pi}(p(K) - p(X))^2 + \pi(\alpha_1^2 + \beta_1^2) + \pi \sum_{n=2}^{\infty} (a_n^2 + b_n^2),$$

$$h_2(K, S(K))^2 = \pi \sum_{n=2}^{\infty} (a_n^2 + b_n^2),$$

$$h_2(S(K), X)^2 = \frac{1}{2\pi}(p(K) - p(X))^2 + \pi(\alpha_1^2 + \beta_1^2).$$

(10) is now an obvious consequence of these three relations.

5. Stability of the Isoperimetric Inequality. We consider now the stability of the isoperimetric inequality $p(K)^2 - 4\pi a(K) \geq 0$ with respect to the deviation measure h_2. As pointed out in the introduction we may formulate stability results as strengthened forms of the original inequality. In the present case we choose the following formulation.

THEOREM 2. *Let K be a convex domain of area $a(K)$ and perimeter $p(K)$, and let $S(K)$ denote the Steiner disc of K. Then,*

$$p(K)^2 - 4\pi a(K) \geq 6\pi h_2(K, S(K))^2. \tag{11}$$

Equality holds if and only if the support function of K is of the form $a_0 + a_1 \cos\theta + a_2 \cos 2\theta + b_1 \sin\theta + b_2 \sin 2\theta$.

Proof. Let us first assume that the smoothness assumptions under which (7), (8), and (9) have been derived are satisfied. Then,

$$\begin{aligned} p(K)^2 - 4\pi a(K) &= 4\pi^2 a_0^2 - 4\pi\left(\pi a_0^2 - \frac{\pi}{2}\sum_{n=2}^{\infty}(n^2-1)\left(a_n^2+b_n^2\right)\right) \\ &= 2\pi^2\sum_{n=2}^{\infty}(n^2-1)\left(a_n^2+b_n^2\right) \geq 6\pi^2\sum_{n=2}^{\infty}\left(a_n^2+b_n^2\right) \\ &= 6\pi h_2(K, S(K))^2. \end{aligned}$$

The general case can be settled either by referring to Blaschke [**1**] where it is shown that (7), (8), and (9) hold without any regularity assumptions, or by using the fact that arbitrary convex domains can be approximated to any desired degree of precision (in the Hausdorff metric) by domains with analytic boundary (cf. [**4**] or [**18**]). Although the remark about the occurrence of the equality sign is an obvious by-product of our proof it has to be noticed that it is only meaningful if there actually are convex domains whose respective support functions are of the indicated type. An example of such a domain can be constructed by considering the curve defined by

$$x = \cos t + \frac{1}{3}\sin^3 t, \quad y = \sin t + \frac{1}{3}\cos^3 t \qquad (0 \leq t \leq 2\pi).$$

It is an easy (but somewhat lengthy) exercise in elementary differential geometry to prove that this is a convex curve whose corresponding convex domain has the support function

$$H(\theta) = 1 + \frac{1}{6}\sin^2\theta.$$

See Inzinger [**23**] for a very detailed study of support functions that are partial sums of trigonometric series.

We add several remarks concerning Theorem 2 and its proof. Our proof of (11) actually yields an explicit expression (in terms of the Fourier coefficients) for the isoperimetric deficit $p(K)^2 - 4\pi a(K)$, namely,

$$p(K)^2 - 4\pi a(K) = 2\pi^2\sum_{n=2}^{\infty}(n^2-1)\left(a_n^2+b_n^2\right). \tag{12}$$

This relationship can be used to obtain a stability statement in terms of the Hausdorff metric. Since it is easily seen that $|a_n\cos n\theta + b_n\sin n\theta| \leq \sqrt{a_n^2+b_n^2}$ it follows that

$$\begin{aligned} \left|H_K(\theta) - H_{S(K)}(\theta)\right| &= \Bigg|a_0 + \sum_{n=1}^{\infty}(a_n\cos n\theta + b_n\sin n\theta) \\ &\qquad - (a_0 + a_1\cos\theta + b_1\sin\theta)\Bigg| \\ &\leq \sum_{n=2}^{\infty}|a_n\cos n\theta + b_n\sin n\theta| \leq \sum_{n=2}^{\infty}\sqrt{a_n^2+b_n^2}. \end{aligned}$$

Hence, using Hölder's inequality, we find

$$h(K, S(K)) \leq \sum_{n=2}^{\infty} \sqrt{a_n^2 + b_n^2} \leq \left(\sum_{n=2}^{\infty} 1/(n^2-1) \right)^{1/2} \left(\sum_{n=2}^{\infty} (n^2-1)(a_n^2 + b_n^2) \right)^{1/2},$$

and because of

$$\sum_{n=2}^{\infty} 1/(n^2-1) = \frac{1}{2} \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n+1} \right) = \frac{3}{4}$$

and (12) we finally obtain

$$p(K)^2 - 4\pi a(K) \geq \frac{8\pi^2}{3} h(K, S(K))^2.$$

But this inequality is not best possible. Better results can be deduced from a theorem of Bonnesen [3] which states that for any K there are two concentric discs of respective radii R and r such that ∂K is contained between these discs and

$$p(K)^2 - 4\pi a(K) \geq 4\pi(R-r)^2. \tag{13}$$

This implies obviously that there is a disc D (of radius $(R+r)/2$, but not necessarily the Steiner disc) such that

$$p(K)^2 - 4\pi a(K) \geq 16\pi h(K, D)^2.$$

Inequalities of the type (13) are known as Bonnesen type inequalities. More information on such inequalities can be found in [**4**] and in the survey [**28**].

Although (11) cannot be improved for all $K \in \mathscr{C}^2$ it is possible to prove stronger inequalities for particular kinds of convex domains. For example, if K is of constant width the Fourier expansion of the support function of K has the property that $a_n = b_n = 0$ for all even values of n. Checking the proof of (11) and the subsequent remarks one sees that this fact leads to the inequalities $p(K)^2 - 4\pi a(K) \geq 16\pi h_2(K, S(K))^2$ and $p(K)^2 - 4\pi a(K) \geq 8\pi^2 h(K, S(K))^2$. From the latter inequality one can derive that $p(K)^2 - 4\pi a(K) \geq 2\pi^2(R-r)^2$, which is an improvement of Bonnesen's inequality in the case of domains of constant width (cf. also [**15**]).

The idea of using Fourier series for proving the isoperimetric inequality is due to Hurwitz [**21**]. With minor modifications this proof has been reproduced in many textbooks (for example in [**7**]). However Hurwitz did not use the Fourier expansion of the support function but employed two series, one for each of the coordinates of the points of the boundary curve. In a subsequent paper [**22**] he gave another proof of the isoperimetric inequality using the Fourier series expansion of the radius of curvature. A proof of the isoperimetric inequality that is based on the support function and Fourier series was first given by Blaschke [**1**] (he used the Fourier series of the derivative of the support function); see also Geppert [**12**]. But none of these articles deals with stability questions. Various other geometric applications of Fourier series have been investigated by Meissner [**26**].

Concerning the stability problem of the isoperimetric inequality in higher dimensions see [**8**], [**29**], and [**32**].

6. Mixed Areas. The isoperimetric inequality can be significantly generalized by the introduction of "mixed areas." These can be defined by the following procedure.

If $K, L \in \mathscr{C}^2$ and α, β are non-negative real numbers we set

$$\alpha K + \beta L = \{\alpha \mathbf{p} + \beta \mathbf{q} : \mathbf{p} \in K, \quad \mathbf{q} \in L\}.$$

One can easily prove that the support function of $\alpha K + \beta L$ is $\alpha H_K + \beta H_L$ (where, of course, H_K and H_L are the support functions of K and L, respectively). Using (6) one finds immediately

$$a(\alpha K + \beta L) = \alpha^2 a(K) + 2\alpha\beta a(K, L) + \beta^2 a(L),$$

with

$$a(K, L) = \frac{1}{2}\int_0^{2\pi} (H_K(\theta) H_L(\theta) - H_K'(\theta) H_L'(\theta))\, d\theta.$$

In particular, $a(K, K) = a(K)$ and, if C is a circular unit disc, $a(K, C) = (1/2)p(K)$. The generalization of the isoperimetric inequality is Minkowski's inequality

$$a(K, L)^2 - a(K)a(L) \geq 0,$$

where equality holds if and only if K and L are homothetic (i.e., $K = \alpha L + \mathbf{p}$ with $\mathbf{p} \in R^n$ and $\alpha > 0$). This inequality has been proved repeatedly, see for example [**1**], [**4**], or [**11**]. The above integral representation of $a(K, L)$ or the obvious relation

$$2a(K, L) = a(K + L) - a(K) - a(L)$$

enable one again to use Fourier series to establish stability results. However, as recently shown by Wallen [**35**], the best approach is probably to first use Fourier series to prove a general inequality, known as Wirtinger's inequality, and then to employ this inequality to show the stability results. Using this method Wallen has proved the following theorem.

Let K and L be two convex domains and let R and r be defined by

$$R = \min\{t: L \subset tK + \mathbf{p}, \quad t > 0, \mathbf{p} \in R^2\},$$

$$r = \max\{s: sK + \mathbf{q} \subset L, \quad s > 0, \mathbf{q} \in R^2\}.$$

Then,

$$a(K, L)^2 - a(K)a(L) \geq \frac{a(K)^2}{4}(R - r)^2. \tag{14}$$

This inequality will fit better into the context of our stability results if it is reformulated in terms of the Hausdorff distance. This yields the following theorem:

THEOREM 3. *If K and L are convex domains, then there is a homothetic copy L' of L such that $a(L') = a(K)$ and*

$$a(K, L)^2 - a(K)a(L) \geq \frac{1}{4}\frac{a(K)a(L)}{d(K)^2} h(K, L')^2, \tag{15}$$

where $d(K)$ denotes the diameter of K.

Proof. In view of (14) we have only to show that there is a homothetic copy L' of L such that $a(L') = a(K)$ and

$$h(K, L') \leq \sqrt{a(K)/a(L)}\,(R - r)\, d(K).$$

Since the case $R = r$ is trivial we may assume $R > r$. The definition of R and r shows that there are points $\mathbf{p}, \mathbf{q} \in R^2$ such that

$$rK + \mathbf{q} \subset L \subset RK + \mathbf{p}.$$

This can also be written in the form

$$r\left(K + \frac{\mathbf{p} - \mathbf{q}}{R - r}\right) \subset L + \left(\frac{r}{R - r}\mathbf{p} - \frac{R}{R - r}\mathbf{q}\right) \subset R\left(K + \frac{\mathbf{p} - \mathbf{q}}{R - r}\right).$$

Hence, replacing (if necessary) K and L by suitable translates we may assume that

$$rK \subset L \subset RK.$$

If we set $\kappa = \sqrt{a(K)/a(L)}$, $L' = \kappa L$, $r' = \kappa r$, and $R' = \kappa R$, then $a(L') = a(K)$ and

$$r'K \subset L' \subset R'K.$$

It follows that $\mathbf{o} \in K$, $\mathbf{o} \in L'$, and $r' \leq 1 \leq R'$. Using these facts we find

$$H_{L'}(\theta) - H_K(\theta) \geq r'H_K(\theta) - H_K(\theta) = -(1 - r')H_K(\theta) \geq -(R' - r')\,d(K)$$

and

$$H_{L'}(\theta) - H_K(\theta) \leq R'H_K(\theta) - H_K(\theta) = (R' - 1)H_K(\theta) \leq (R' - r')\,d(K).$$

Consequently we obtain the desired result

$$h(K, L') \leq (R' - r')\,d(K) = \sqrt{a(K)/a(L)}\,(R - r)\,d(K).$$

A kind of n-dimensional generalization of Theorem 3 has been given by Diskant [**8**]. However in the case $n = 2$ the result of Diskant is weaker than (15) (see also Schneider [**32**]).

7. The Brunn-Minkowski Theorem. The Brunn-Minkowski theorem is one of the most important theorems in the theory of convex sets. In the case $n = 2$ it is the following statement concerning the area of the linear combination $\alpha K + (1 - \alpha)L$ of two convex domains K and L.

If $K \in \mathscr{C}^2$, $L \in \mathscr{C}^2$, *and* $0 \leq \alpha \leq 1$, *then*

$$a(\alpha K + (1 - \alpha)L)^{1/2} \geq \alpha a(K)^{1/2} + (1 - \alpha)a(L)^{1/2} \tag{16}$$

In the case $0 < \alpha < 1$ *equality holds if and only if* K *and* L *are homothetic.*

For the sake of simplicity let us consider the special situation when $a(K) = a(L) = 1$. (The general case can be derived from this restricted version by simple substitutions.) Under this assumption (16) can be written as

$$a(\alpha K + (1 - \alpha)L) \geq 1.$$

The stability problem associated with this inequality is to estimate the degree of homothetic relationship of K and L if for some α $(0 < \alpha < 1)$ and $\varepsilon \geq 0$

$$a(\alpha K + (1 - \alpha)L) \leq 1 + \varepsilon.$$

In [**14**] where the n-dimensional case is considered it is shown that this inequality implies the existence of a translate L' of L such that

$$h(K, L') \leq 4\sqrt{3}\left(\frac{1}{\sqrt{\alpha(1 - \alpha)}} + 2\right) D\varepsilon^{1/3},$$

where $D = \max\{d(K), d(L)\}$. Using Theorem 3 we can here easily prove the following stronger result.

THEOREM 4. *Let K and L be two convex domains with $a(K) = a(L) = 1$, and let α be such that $0 < \alpha < 1$. If for some $\varepsilon \geq 0$*

$$a(\alpha K + (1 - \alpha)L) \leq 1 + \varepsilon, \tag{17}$$

then there exists a translate L' of L such that

$$h(K, L') \leq \frac{\sqrt{17}}{2} \frac{1}{\sqrt{\alpha(1 - \alpha)}} D\varepsilon^{1/2}, \tag{18}$$

where $D = \max\{d(K), d(L)\}$.

Proof. We may assume that

$$\sqrt{17\varepsilon}/2\sqrt{\alpha(1 - \alpha)} \leq 1, \tag{19}$$

since if this is not true then (18) is trivially satisfied. Using the definition of $a(K, L)$ we see that (17) implies

$$(1 - \alpha)^2 + 2\alpha(1 - \alpha)a(K, L) + \alpha^2 \leq 1 + \varepsilon.$$

This can also be written as

$$a(K, L) - 1 \leq \varepsilon/(2\alpha(1 - \alpha)).$$

Consequently, using also (19), we find

$$a(K, L)^2 - 1 \leq \frac{\varepsilon}{2\alpha(1 - \alpha)}(1 + a(K, L)) \leq \frac{\varepsilon}{2\alpha(1 - \alpha)}\left(2 + \frac{2}{17}\right)$$

$$= \frac{\varepsilon}{\alpha(1 - \alpha)} \frac{18}{17}.$$

Hence, Theorem 3 implies that there is a translate L' of L such that

$$h(K, L')^2 \leq \frac{\varepsilon}{\alpha(1 - \alpha)} \frac{18}{17} 4D^2 \leq \frac{17}{4} \frac{\varepsilon}{\alpha(1 - \alpha)} D^2,$$

and (18) follows from this.

As an illustration of the use of Theorem 4 we mention a stability result concerning central symmetrization. If $K \in \mathcal{C}^2$, we set $-K = \{-p: p \in K\}$ and call the set

$$K^* = \frac{1}{2}(K + (-K)),$$

the convex domain obtained from K by central symmetrization. K^* is obviously a centrally symmetric convex domain and the Brunn-Minkowski theorem shows that

$$a(K^*) \geq a(K)$$

with equality if and only if K itself is centrally symmetric. Thus we meet the stability problem of deciding how much K deviates from a centrally symmetric convex domain if $a(K^*) \leq a(K) + \varepsilon$. The following corollary addresses this problem.

COROLLARY. *Let K be a convex domain of area 1. If $a(K^*) \leq 1 + \varepsilon$ (for some $\varepsilon \geq 0$), then there is a centrally symmetric convex domain Z (which is a translate of K^*) such that*

$$h(K, Z) \leq \frac{1}{2}\sqrt{17\varepsilon}\, d(K).$$

Proof. If Theorem 4 is applied to K and $-K$ we obtain a translate $(-K)'$ of $-K$ such that $h(K,(-K)') \leq \sqrt{17\varepsilon}\, d(K)$. If we set $Z = (1/2)(K + (-K)')$, then Z is a translate of K^* (and therefore centrally symmetric) and

$$\begin{aligned} |H_Z(\theta) - H_K(\theta)| &= \left|\frac{1}{2}\big(H_K(\theta) + H_{(-K)'}(\theta)\big) - H_K(\theta)\right| \\ &= \frac{1}{2}|H_K(\theta) - H_{(-K)'}(\theta)| \leq \frac{1}{2}h(K,(-K)'). \end{aligned}$$

Hence, $h(K, Z) \leq (1/2)h(K,(-K)')$ and the desired result follows.

This Corollary has applications in the theory of packings of convex domains. It enables one to estimate how much a convex domain K deviates from a centrally symmetric convex domain if K permits packings (by translates) of high density. See [**16**] for more details. An n-dimensional version of the Corollary together with an application, but with poorer constants in the case $n = 2$, is given in [**14**].

There is another kind of stability problem that can be associated with the Brunn-Minkowski theorem. Instead of assuming that (16) holds for a particular value α one may suppose that it holds for all α. Diskant [**9**] has investigated the stability of the Brunn-Minkowski theorem from this point of view. However, in most applications, like that concerning central symmetrization, the available information is restricted to only one value of α.

9. Stability and Projections. To indicate the wide variety of stability problems in the theory of convex sets we discuss now a problem that is not exactly of the type as described in the Introduction. If K is a convex body and E a plane in R^3 let K_E denote the orthogonal projection of K onto E. Assume now that every projection K_E differs (with respect to some deviation measure) at most ε from a circular disc. What can be said about the deviation of K from a ball? To give a precise formulation of this problem let $R(M)$ denote the circumradius and $r(M)$ the inradius of a convex body M in R^2 or R^3. Among much more general results dealing with this problem the following theorem is proved in [**13**]:

Assume $K \in \mathcal{C}^3$ and $\varepsilon \geq 0$. If for every plane E in R^3

$$R(K_E) - r(K_E) \leq \varepsilon, \tag{20}$$

then

$$R(K) - r(K) \leq 6(1 + 2\sqrt{2})\varepsilon. \tag{21}$$

We can use this theorem to prove the following result which, to some extent, answers a question of Chakerian [**6**]. The question is whether one can conclude that the isoperimetric quotient of a convex body in R^3 is large (i.e., close to 1) if every projection has a large isoperimetric quotient. In our context it is more convenient to work with the isoperimetric deficits $p(K_E)^2 - 4\pi a(K_E)$ and $\sigma(K)^{1/2} - (36\pi)^{1/6}v(K)^{1/3}$, where σ denotes surface area and v volume.

THEOREM 5. *Let there be given a convex body K in R^3 and an $\varepsilon \geq 0$. If for every plane E in R^3*

$$p(K_E)^2 - 4\pi a(K_E) \leq \varepsilon, \tag{22}$$

then

$$\sigma(K)^{1/2} - (36\pi)^{1/6} v(K)^{1/3} \leq 6(1 + 2\sqrt{2})\sqrt{\varepsilon}\,.$$

Proof. From (22) and Bonnesen's inequality (13) it follows that

$$R(K_E) - r(K_E) \leq \frac{1}{\sqrt{4\pi}} E.$$

Since (20) implies (21) we obtain from this inequality

$$R(K) - r(K) \leq (1 + 2\sqrt{2})\frac{3}{\sqrt{\pi}} E.$$

Using this we find

$$\begin{aligned}\sigma(K)^{1/2} - (36\pi)^{1/6} v(K)^{1/3} &\leq \left(4\pi R(K)^2\right)^{1/2} - (36\pi)^{1/6}\left(\frac{4}{3}\pi r(K)^3\right)^{1/3} \\ &= \sqrt{4\pi}\,(R(K) - r(K)) \leq 6(1 + 2\sqrt{2})\sqrt{\varepsilon}\,.\end{aligned}$$

10. Concluding Remarks. There are many other theorems in the theory of convex sets that give rise to interesting stability problems. Actually most classical characterizations of simplices, balls, ellipsoids etc. have significant associated stability problems. In some cases attention has to be restricted to much smaller classes than the class of all convex bodies. For example, it is known that a convex domain of given constant width and minimal area is a Reuleaux triangle. The associated stability problem is to determine how much a convex domain of constant width deviates from a Reuleaux triangle if its area is close to the minimal area. An answer to this problem is given in [**15**] and the analogous problem for equichordal sets is considered in [**17**]. Some of the stability results that have been proved by the use of Fourier series can be generalized to the n-dimensional situation if instead of Fourier series one uses expansions in terms of spherical harmonics (see, for example, the recent work of Schneider [**32**]).

REFERENCES

1. Blaschke, W., Beweise zu Sätzen von Brunn-Minkowski über die Minimaleigenschaft des Kreises, *Jahresber. Deutsch. Math.-Verein.*, 23 (1914) 210–234.
2. Blaschke, W., Kreis und Kugel, Walter de Gruyter & Co., Berlin 1956.
3. Bonnesen, T., Problèmes des Isopérimètres et de Isépiphanes. Gauthier-Villars et Cie., Paris, 1929.
4. Bonnesen, T. and Fenchel, W., Theorie der konvexen Körper. Ergebn. d. Math. Bd. 3, J. Springer Verlag, Berlin 1934.
5. Campi, S., Reconstructing a convex surface from certain measurements of its projection, *Boll. Un. Mat. Ital.* (6) 5-B (1986) 945–959.
6. Chakerian, G. D., Is a body spherical if all its projections have the same I.Q.?, *Amer. Math. Monthly*, 77 (1970) 989–992.
7. Courant, R. and Hilbert, D., Methods of Mathematical Physics., Vol. 1, Wiley (Interscience), New York, 1953.
8. Diskant, V. I., Bounds for the discrepancy between convex bodies in terms of the isoperimetric deficit, *Sibirsk. Mat. Ž.*, 13 (1972) 767–772. (Transl. in: *Sibirian Math. J.*, 13 (1972) 529–532.)

9. _____, Stability of the solution of the Minkowski equation, *Sibirsk. Math. Ž.*, 14 (1973) 669–673, 696. (Transl. in *Sibirian Math. J.*, 14 (1973) 466–469.)
10. Eggleston, H. G., Convexity, Cambr. Tracts in Math. and Math. Phys. No. 47, Cambridge Univ. Press, New York 1959.
11. Flanders, H., A proof of Minkowski's inequality for convex curves, *Amer. Math. Monthly*, 75 (1968) 581–593.
12. Geppert, H., Über den Brunn-Minkowskischen Satz, *Math. Z.*, 42 (1937) 238–254.
13. Groemer, H., Stability theorems for projections of convex sets, *Israel J. Math.*, 60 (1987) 177–190.
14. _____, On the Brunn-Minkowski Theorem, *Geom. Dedicata*, 27 (1988) 357–371.
15. _____, Stability theorems for convex domains of constant width, *Canad. Math. Bull.*, 31 (1988) 328–337.
16. _____, On convex bodies that permit packings of high density, *Discrete Comput. Geom.* (in press).
17. _____, On the volume of equichordal sets, *Monatsh. Math.*, 106 (1988) 1–8.
18. Gruber, P. M., Approximation of convex bodies. In: Convexity and Its Applications, P. M. Gruber and J. Wills (editors), Birkhäuser Verlag, Basel-Boston-Stuttgart, 1983, pp. 131–162.
19. Grünbaum, B., Convex Polytopes, John Wiley and Sons, London-New York-Sydney, 1967.
20. Hadwiger, H., Altes und Neues über konvexe Körper, Birkhäuser Verlag, Basel-Stuttgart, 1955.
21. Hurwitz, A., Sur le problème de isopérimetres, *C. R. Acad. Sci. Paris*, 132 (1901) 401–403. (Math. Werke, 1. Bd., Birkhäuser, Basel, 1932, pp. 490–491.)
22. Hurwitz, A., Sur quelque applications géométriques des séries de Fourier, *Ann. Ecole Normal Sup.* (3) 19 (1902) 357–408. (Math. Werke, 1. Bd., Birkhäuser, Basel, 1932, pp. 509–554.)
23. Inzinger, R., Stützbare Bereiche, trigonometrische Polynome und Defizite höherer Ordnung, *Monatsh. Math.*, 53 (1949) 302–323.
24. Kubota, T., Über die Schwerpunkte der konvexen geschlossenen Kurven und Flächen, *Tahoku Math. J.*, 14 (1918(20–27.)
25. Leichtweiss, K., Konvexe Mengen. Springer Verlag, Berlin-Heidelberg-New York, 1980.
26. Meissner, E., Über die Anwendung von Fourier-Reihen auf einige Aufgaben der Geometrie und Kinematik. *Vierteljahresschrd. naturforschenden Gesellsch. Zürich*, 54 (1909) 309–329.
27. Minkowski, H., Über die Körper konstanter Breite. Ges. Abh., 2. Bd., pp. 277–279, Teubner, Leipzig-Berlin, 1911.
28. Osserman, R., Bonnesen-Style isoperimetric inequalities, *Amer. Math. Monthly*, 86 (1979) 1–29.
29. _____, A strong form of the isoperimetric inequality in R^n, *Complex Variables Theory Appl.*, 9 (1987) 241–49.
30. Saint-Pierre, J., Point de Steiner et section lipschitziennes, *Sem. Anal. Convexe* 15 (1985) Exp. No. 7; p. 42.
31. Schneider, R., Krümmungsschwerpunkte konvexer Körper I, II, *Abh. Math. Sem. Univ. Hamburg*, 37 (1972) 112–132, 204–217.
32. Schneider, R., Stability in the Aleksandrov-Fenchel-Jessen theorem (to appear in *Mathematica*).
33. Shephard, G. C. and Webster, R. J., Metrics for sets of convex bodies, *Mathematika*, 12 (1965) 73–88.
34. Vitale, R. A., L_p metrics for compact convex sets, *J. Approx. Theory*, 45 (1985) 280–287.
35. Wallen, L. J., All the way with Wirtinger: A short proof of Bonnesen's inequality, *Amer. Math. Monthly*, 94 (1987) 440–442.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

Separable Extensions and Diagonalizability

FRED RICHMAN
Department of Mathematics, New Mexico State University, Las Cruces, NM 88003

1. Introduction. The notion of a separable field extension plays a major role in many treatments of Galois theory. As any algebraic extension of a field of characteristic 0 is separable, expositions of this notion tend to focus on fields of characteristic p. The idea appears even more esoteric because algebraic extensions of finite fields are also separable. Thus in a subject that grows out of the study of algebraic numbers, we must use rational function fields of characteristic p to give content to one of the central concepts.

By enlarging the scope of the discussion to include extensions with zero-divisors we can construct natural examples of inseparability in characteristic 0, we get to observe the connection between separability and diagonalizability of matrices, and we don't have to mention the characteristic of the field. As an added bonus, our theorems are more general.

In this note we establish, in this enlarged context, the fundamental results that the separable elements in a finite-dimensional commutative extension of a field k form a ring, and, if k is infinite, the ring is of the form $k[\theta]$. The basic techniques, which are not new, are diagonalizing a matrix whose minimal polynomial has no repeated roots, and simultaneously diagonalizing commuting matrices.

2. Separability and diagonalizability. A polynomial p over a field k is said to be *separable* if it is relatively prime to its formal derivative, that is, if there exist polynomials s and t such that $sp + tp' = 1$. The polynomials s and t can be constructed so that their coefficients are in the field generated by the coefficients of p; indeed the Euclidean algorithm does just that. It is easy to verify that the polynomial $(X - r_1)(X - r_2) \cdots (X - r_n)$ is separable if and only if all the r_i are distinct. Thus a polynomial is separable if and only if it has no repeated roots in any extension field, so any factor of a separable polynomial is separable.

Many authors define a polynomial to be separable if it is a product of polynomials that are separable in the above sense. This makes the separability of p dependent on the field k: for example, any polynomial is then separable over an algebraically closed field. For irreducible polynomials it is all the same, so for minimal polynomials of elements in an extension field of k, we can use either definition. However, if we allow zero-divisors in the extension of k, then the minimal polynomials need not be irreducible, and the two definitions are quite different.

By the *field of definition* of a finite set of matrices over a field, we mean the (countable) field generated by their entries. If matrices are linearly dependent over a field k, then they are already linearly dependent over their field of definition:

indeed if we treat the matrices as if they were row vectors, then the usual procedure of Gaussian elimination will either result in a nontrivial dependence relation over the field of definition, or will demonstrate that the rows are independent over k. In particular, the coefficients of the minimal polynomial of a matrix lie in the field of definition of that matrix, so the minimal polynomial of a matrix does not depend on the particular field from which the entries of the matrix are considered to be taken. Also, if a matrix A can be written as a polynomial in a matrix B, then the polynomial may be chosen with coefficients in the field of definition of A and B.

An n-by-n matrix A with entries in a field k is diagonalizable over k if and only if k^n is the direct sum of the eigenspaces of A, that is, if k^n admits a basis of eigenvectors of A. The connection between separable polynomials and diagonalizable matrices is

THEOREM 1. *Let A be a matrix with entries in a field. Then the minimal polynomial of A is separable if and only if A is diagonalizable over some field k. In this case, the projections of k^n onto the eigenspaces of A can be written as polynomials in A.*

Proof. Suppose the matrix A is diagonalizable over a field k, and let Λ be the set of eigenvalues of A. Then A satisfies the separable polynomial $\Pi_{\lambda \in \Lambda}(X - \lambda)$. Conversely, if A satisfies a separable polynomial, let k be a splitting field for the minimal polynomial $f(X)$ of A over the field of definition of A. Write $f(X) = (X - \lambda)g_\lambda(X)$ for each root λ of f.

As $(A - \lambda)g_\lambda(A) = f(A) = 0$, the matrix $g_\lambda(A)$ maps into V_λ, the λ-eigenspace of A. If $\mu \neq \lambda$ are in Λ, then $A - \mu$ is a factor of $g_\lambda(A)$, so $g_\lambda(A)V_\mu = 0$. Finally, the polynomial $\Sigma_{\lambda \in \Lambda} g_\lambda(X)/g_\lambda(\lambda)$ takes on the value 1 for each λ in Λ, and has degree less than the cardinality of Λ, so is identically equal to 1. Thus $\Sigma_{\lambda \in \Lambda} g_\lambda(A)/g_\lambda(\lambda)$ is the identity matrix, so $g_\lambda(A)/g_\lambda(\lambda)$ is the projection of k^n onto V_λ. □

The last sentence of Theorem 1 says that if A is a diagonal matrix, and P_λ is the diagonal matrix with 1's where A has λ's, and 0's elsewhere, then P_λ can be written as a polynomial in A. This fact will play a key role in proving the existence of primitive elements (Theorem 5).

3. Simultaneous diagonalizability and primitive elements. The standard result regarding simultaneous diagonalizability of commuting matrices is

THEOREM 2. *Let A and B be commuting n-by-n matrices with entries in a field k. If k^n admits a basis of eigenvectors for each of A and B, then k^n admits a basis whose elements are eigenvectors of both A and B.*

Proof. Let V_λ^A and V_λ^B be the λ-eigenspace of A and B respectively. Then $k^n = \oplus_\lambda V_\lambda^A = \oplus_\lambda V_\lambda^B$ by hypothesis. As B commutes with $A - \lambda$, the subspace V_λ^A is invariant under B, hence also under the projection π_μ^B of k^n onto V_μ^B, which is a polynomial in B. Therefore $V_\lambda^A = \Sigma_\mu V_\lambda^A \cap V_\mu^B$, so $k^n = \Sigma_{\lambda,\mu} V_\lambda^A \cap V_\mu^B$ □

If A and B are elements of a commutative ring containing k, then we denote by $k[A, B]$ the subring of elements that can be written as polynomials in A and B with coefficients in k.

COROLLARY 3. *If A and B are commuting separable matrices, then any matrix, or inverse of a matrix, in $k[A, B]$ is separable.*

Proof. By Theorems 1 and 2 we can simultaneously diagonalize A and B. Then all the matrices in question are diagonalizable, hence, separable by Theorem 1. □

Results about matrices over k can be extended to elements in arbitrary finite-dimensional extension rings K of k through the device of passing from an element α in K to the matrix of the linear transformation from K to K induced by multiplication by α. An element of a commutative ring extension of a field k is said to be *separable* over k if it satisfies a separable polynomial over k, or, equivalently, if its minimal polynomial is separable.

COROLLARY 4. *If K is a commutative ring containing a field k, then the elements of K that are separable over k form a ring.*

Proof. It suffices to show that if α and β are separable over k, then every element of the ring $k[\alpha, \beta]$ is separable over k. The ring $k[\alpha, \beta]$ is a finite-dimensional vector space over k as it is generated by the monomials $\alpha^i\beta^j$ where i and j are less than the degrees of the minimal polynomials of α and β respectively. Let T_α and T_β be the linear transformations given by multiplication by α and β on the finite-dimensional vector space $k[\alpha, \beta]$. Observe that $k[T_\alpha, T_\beta]$ is naturally isomorphic to $k[\alpha, \beta]$. Let A and B be the matrices of T_α and T_β with respect to some basis of $k[\alpha, \beta]$. By Corollary 3 any matrix in $k[A, B] \cong k[T_\alpha, T_\beta\]$ is separable, hence any element of $k[\alpha, \beta]$ is separable. □

Proofs of the existence of a primitive element for finite-dimensional separable extensions K of k, that is, an element θ such that $K = k[\theta]$, divide into two cases depending on whether k is finite or infinite. The finite case follows from the fact that K is also finite, so we can choose θ to be a generator of the multiplicative group of K. If we allow K to have zero-divisors, then there need not be a primitive element in the finite case: for example, let K be the ring of 3-by-3 diagonal matrices over the two-element field. The infinite case, however, goes through in complete generality. By induction it suffices to show that $k[A, B] = k[C]$ where A and B are commuting separable matrices.

THEOREM 5. *If A and B are commuting separable n-by-n matrices over a field k of cardinality greater than $n(n-1)/2$, then there exists an element c in k such that A and B can be written as polynomials in $A + cB$ with coefficients in k.*

Proof. We may assume that A and B are diagonal with diagonal elements $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$. Choose c distinct from $(a_j - a_i)/(b_i - b_j)$ for each pair i and j such that $b_i \neq b_j$. Then $a_i + cb_i \neq a_j + cb_j$ whenever $a_i \neq a_j$ or $b_i \neq b_j$. It follows from the last sentence of Theorem 1 that A and B can be written as polynomials in $A + cB$. □

If all we want is a matrix C such that $k[A, B] = k[C]$, the lower bound $n(n-1)/2$ for the field in Theorem 5 can be replaced by $n - 1$; this bound is sharp in that, over a field of cardinality $n - 1$, all the n-by-n diagonal matrices can be written as polynomials in two diagonal matrices, but not in one diagonal matrix. To establish the bound $n - 1$, we may assume, as before, that A and B are diagonal with diagonal elements $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$. It suffices to construct C in $k[A, B]$ whose diagonal entries differ whenever the corresponding entries of A or B differ, because then we can write A and B as polynomials in C by the Lagrange

interpolation formula. By Theorem 1 the diagonal matrix E_t whose mth entry is 1 if $(a_m, b_m) = (a_t, b_t)$, and 0 otherwise, is in $k[A, B]$ for each i. Note that either $E_i = E_j$ or $E_iE_j = 0$, and the sum of the distinct E_t is the identity. As k has at least n elements, we can let C be a linear combination of the distinct E_t with distinct coefficients from k.

The hypotheses of Theorem 5 can be weakened so that only A is assumed to be separable over k; here is a brief sketch of how to do that. Put A in block diagonal form, $\operatorname{diag}(A_1, \ldots, A_m)$, where each block A_t is a scalar multiple of an identity matrix, and the scalars are distinct in different blocks. As B commutes with A, we must have $B = \operatorname{diag}(B_1, \ldots, B_m)$, where the block B_t is of the same size as the block A_t. Then put each B_t in Jordan canonical form; this does not affect the blocks A_t because A_t commutes with everything. Now choose c, by Theorem 5, so that A and the diagonal of B can be written as polynomials in the diagonal of $A + cB$. It remains to show that the diagonal of $A + cB$ can be written as a polynomial in $A + cB$. This is true for any block triangular matrix M in which each block has a constant diagonal: if the minimal polynomial $f(X)$ of M is of the form $(X - \lambda)^e g(X)$, and $s(X)(X - \lambda)^e + t(X)g(X) = 1$, then the matrix $t(A)g(A)$ is diagonal, with 1's where M has λ's, and 0's elsewhere; the diagonal of M is a linear combination of these matrices.

A Birthday Present for Ramanujan

M. D. HIRSCHHORN
University of New South Wales, Kensington, NSW, Australia 2033

Recently, Frank G. Garvan and George E. Andrews made an exciting discovery. Given a partition π, let $\lambda(\pi)$ denote the largest part of π, $\mu(\pi)$ denote the number of ones in π and $\nu(\pi)$ denote the number of parts of π larger than $\mu(\pi)$. Define the "crank" of the partition as follows:

$$c(\pi) = \begin{cases} \lambda(\pi) & \text{if } \mu(\pi) = 0 \\ \nu(\pi) - \mu(\pi) & \text{if } \mu(\pi) > 0. \end{cases}$$

Then the partitions of $5n + 4$ split into 5 equinumerous classes according to the residue modulo 5 of the crank, the partitions of $7n + 5$ split into 7 equinumerous classes according to the residue modulo 7 of the crank, and the partitions of $11n + 6$ split into 11 equinumerous classes according to the residue modulo 11 of the crank. Thus Garvan and Andrews have succeeded in giving a combinatorial explanation of the congruences due to Ramanujan, $p(5n + 4) \equiv 0 \mod 5$, $p(7n + 5) \equiv 0 \mod 7$ and $p(11n + 6) \equiv 0 \mod 11$.

The significance of this discovery is that the existence of such a "crank" was postulated more than forty years ago by Freeman Dyson, but it has remained hidden for so long.

The proof that the crank does what is claimed comes in two parts. First it is shown that the crank is generated (essentially) by the infinite product

$$\frac{(q;q)_\infty}{(a^{-1}q;q)_\infty(aq;q)_\infty}\cdot\left[(a;q)_\infty=\prod_{n\geq 0}(1-aq^n)\right].$$

The second part of the proof consists in showing that when a is replaced by a fifth, seventh or eleventh root of unity and the product is expanded, certain powers of q do *not* appear.

Thus if η is a fifth root of unity, we have the identity, to be found in the so-called lost notebook of Ramanujan,

$$\begin{aligned}\frac{(q;q)_\infty}{(\eta^{-1}q,q)_\infty(\eta q;q)_\infty}=\frac{1}{(q^5;q^5)_\infty}\Bigg\{&\frac{(q^{10};q^{25})_\infty^2(q^{15};q^{25})_\infty^2(q^{25};q^{25})_\infty^2}{(q^5;q^{25})_\infty(q^{20};q^{25})_\infty}\\&+(\eta-1+\eta^{-1})q(q^{10};q^{25})_\infty(q^{15};q^{25})_\infty(q^{25};q^{25})_\infty^2\\&-(\eta+1+\eta^{-1})q^2(q^5;q^{25})_\infty(q^{20};q^{25})_\infty(q^{25};q^{25})_\infty^2\\&-(\eta+\eta^{-1})q^3\frac{(q^5;q^{25})_\infty^2(q^{20};q^{25})_\infty^2(q^{25};q^{25})_\infty^2}{(q^{10};q^{25})_\infty(q^{15};q^{25})_\infty}\Bigg\}.\end{aligned}$$

That no powers of q congruent to 4 modulo 5 occur on the right of this identity provides the proof that the partitions of $5n+4$ split into 5 equinumerous classes as claimed earlier.

Frank Garvan has found the corresponding identity for a seventh root of unity, in which no powers of q congruent to 5 modulo 7 appear, namely,

$$\begin{aligned}\frac{(q;q)_\infty}{(\eta^{-1}q;q)_\infty(\eta q;q)_\infty}=\frac{1}{(q^7;q^7)_\infty}\Big\{&(q^{21};q^{49})_\infty^2(q^{28};q^{49})_\infty^2(q^{49};q^{49})_\infty^2\\&+(\eta-1+\eta^{-1})q(q^{14};q^{49})_\infty(q^{21};q^{49})_\infty(q^{28};q^{49})_\infty(q^{35};q^{49})_\infty(q^{49};q^{49})_\infty^2\\&+(\eta^2+\eta^{-2})q^2(q^{14};q^{49})_\infty^2(q^{35};q^{49})_\infty^2(q^{49};q^{49})_\infty^2\\&-(\eta^2+\eta+\eta^{-1}+\eta^{-2})q^3(q^7;q^{49})_\infty(q^{21};q^{49})_\infty(q^{28};q^{49})_\infty(q^{42};q^{49})_\infty(q^{49};q^{49})_\infty^2\\&-(\eta+\eta^{-1})q^4(q^7;q^{49})_\infty(q^{14};q^{49})_\infty(q^{35};q^{49})_\infty(q^{42};q^{49})_\infty(q^{49};q^{49})_\infty^2\\&-(\eta^2+1+\eta^{-2})q^6(q^7;q^{49})_\infty^2(q^{42};q^{49})_\infty^2(q^{49};q^{49})_\infty^2\Big\}.\end{aligned}$$

Further, he has shown that in the corresponding identity involving an eleventh root of unity no powers of q congruent to 6 modulo 11 appear, thereby completing the proof.

I have found a simpler form than Garvan's of this last identity, and would like to present it here, on the occasion of the centenary of Ramanujan's birth, in memory of the man whose life and work have been an inspiration to me and to so many others.

The identity is

$$\frac{(q;q)_\infty}{(\eta^{-1}q;q)_\infty(\eta q;q)_\infty}$$

$$= \frac{1}{(q^{11};q^{11})_\infty} \times \Big\{ (q^{22};q^{121})_\infty (q^{33};q^{121})_\infty (q^{44};q^{121})_\infty (q^{55};q^{121})_\infty$$

$$\cdot (q^{66};q^{121})_\infty (q^{77};q^{121})_\infty (q^{88};q^{121})_\infty (q^{99};q^{121})_\infty (q^{121};q^{121})_\infty^2$$

$$+ (\eta - 1 + \eta^{-1}) q (q^{11};q^{121})_\infty (q^{44};q^{121})_\infty (q^{55};q^{121})_\infty^2$$

$$\cdot (q^{66};q^{121})_\infty^2 (q^{77};q^{121})_\infty (q^{110};q^{121})_\infty (q^{121};q^{121})_\infty^2$$

$$+ (\eta^2 + \eta^{-2}) q^2 (q^{22};q^{121})_\infty (q^{33};q^{121})_\infty^2 (q^{55};q^{121})_\infty$$

$$\cdot (q^{66};q^{121})_\infty (q^{88};q^{121})_\infty^2 (q^{99};q^{121})_\infty (q^{121};q^{121})_\infty^2$$

$$+ (\eta^3 + 1 + \eta^{-3}) q^3 (q^{22};q^{121})_\infty^2 (q^{44};q^{121})_\infty (q^{55};q^{121})_\infty$$

$$\cdot (q^{66};q^{121})_\infty (q^{77};q^{121})_\infty (q^{99};q^{121})_\infty^2 (q^{121};q^{121})_\infty^2$$

$$+ (\eta^4 + \eta^2 + 1 + \eta^{-2} + \eta^{-4}) q^4 (q^{11};q^{121})_\infty (q^{33};q^{121})_\infty (q^{44};q^{121})_\infty (q^{55};q^{121})_\infty$$

$$\cdot (q^{66};q^{121})_\infty (q^{77};q^{121})_\infty (q^{88};q^{121})_\infty (q^{110};q^{121})_\infty (q^{121};q^{121})_\infty^2$$

$$- (\eta^4 + \eta^2 + \eta^{-2} + \eta^{-4}) q^5 (q^{11};q^{121})_\infty (q^{33};q^{121})_\infty (q^{44};q^{121})_\infty^2$$

$$\cdot (q^{77};q^{121})_\infty^2 (q^{88};q^{121})_\infty (q^{110};q^{121})_\infty (q^{121};q^{121})_\infty^2$$

$$+ (\eta^4 + \eta + \eta^{-1} + \eta^{-4}) q^7 (q^{11};q^{121})_\infty (q^{22};q^{121})_\infty (q^{44};q^{121})_\infty (q^{55};q^{121})_\infty$$

$$\cdot (q^{66};q^{121})_\infty (q^{77};q^{121})_\infty (q^{99};q^{121})_\infty (q^{110};q^{121})_\infty (q^{121};q^{121})_\infty^2$$

$$+ (\eta^4 + \eta^3 + \eta + \eta^{-1} + \eta^{-3} + \eta^{-4}) q^{19} (q^{11};q^{121})_\infty^2 (q^{22};q^{121})_\infty (q^{33};q^{121})_\infty$$

$$\cdot (q^{88};q^{121})_\infty (q^{99};q^{121})_\infty (q^{110};q^{121})_\infty^2 (q^{121};q^{121})_\infty^2$$

$$- (\eta^4 + 1 + \eta^{-4}) q^9 (q^{11};q^{121})_\infty (q^{22};q^{121})_\infty (q^{33};q^{121})_\infty (q^{55};q^{121})_\infty$$

$$\cdot (q^{66};q^{121})_\infty (q^{88};q^{121})_\infty (q^{99};q^{121})_\infty (q^{110};q^{121})_\infty (q^{121};q^{121})_\infty^2$$

$$- (\eta^3 + \eta^{-3}) q^{10} (q^{11};q^{121})_\infty (q^{22};q^{121})_\infty (q^{33};q^{121})_\infty (q^{44};q^{121})_\infty$$

$$\cdot (q^{77};q^{121})_\infty (q^{88};q^{121})_\infty (q^{99};q^{121})_\infty (q^{110};q^{121})_\infty (q^{121};q^{121})_\infty^2 \Big\}.$$

A Condition for Flatness of Curves in R^n

YAAKOV S. KUPITZ AND MICHA A. PERLES
Institute of Mathematics, Hebrew University of Jerusalem, Givat Ram, 91904 Jerusalem

Introduction. The following two exercises are standard in a first course on the differential geometry of curves, and appear in this form or another in many textbooks. (See, e.g., [**1**, Ex. 8 on p. 97] or [**2**, Ex. 4.10 on p. 74 and Ex. 4.20 on p. 77].)

1. Prove that if all tangent lines to a curve pass through a fixed point, then the curve lies on a straight line.

2. Prove that if all the osculating planes to a curve pass through a fixed point, then the curve lies in a plane. (An important special case: the path of a free particle in a central force field is planar; for example, planetary motion.)

The connection between these two problems is obvious, and they just "call" for a common generalization. Unfortunately, this call is ignored in the standard textbooks. In this note we present such a generalization: "If all the osculating k-flats to a curve in R^n (k fixed, $1 \leq k < n$) pass through one point, then the curve lies in a k-flat."

For $k = 1$ and $k = 2$, this generalization reduces to the assertions of exercises 1 and 2.

Another point left obscure in the textbooks is: what smoothness conditions are needed in exercises 1 and 2? The wording of the exercises and their solutions in [**2**] seem to indicate that the curves in question should be of class C^{k+1} (for $k = 1$ and $k = 2$, respectively). At any rate, since the definition of the osculating k-flat (see §1) involves the first k derivatives of the curve at a point, it is natural to assume (at least) that the curve is k times differentiable. We shall, in fact, prove the generalization under exactly this "minimal" smoothness condition, not assuming even the continuity of the kth derivative. See, however, §5 for a possible attempt to relax the smoothness conditions even further.

For a good introduction to curves in R^n we refer the reader to [**3**].

1. Statement of results. Consider a curve in R^n, $n \geq 2$, i.e., a continuous map $X: I \to R^n$. Here I is a real interval, which may be bounded or unbounded, with or without endpoints. For $J \subset I$, put $X(J) = \{X(t): t \in J\}$. For fixed k, $1 \leq k < n$, we shall seek conditions to ensure that $X(I)$ is included in a k-flat, i.e., in a k-dimensional affine subspace of R^n. Since the conditions that we are going to impose pertain to the behavior of X at interior points of I only, and any k-flat that includes $X(\text{int}\, I)$ includes $X(I)$, we shall assume, without loss of generality, that I is open.

DEFINITION 1. *The curve $X: I \to R^n$ is a k-curve if*:

(a) *X is k times differentiable*; (*We denote by $X^{(i)}(t)$ the ith derivative of X at t.*) *and*

(b) *for each $t \in I$, the k vectors $X^{(1)}(t), X^{(2)}(t), \ldots, X^{(k)}(t)$ are linearly independent.* (*In particular, $X^{(1)}(t) \neq 0$ for all $t \in I$.*)

Note that we do not assume continuity of the kth derivative.

Define also, for $t \in I$:

$$L_0(t) = \operatorname{span}\{X^{(1)}(t), \ldots, X^{(k)}(t)\},$$
$$L(t) = X(t) + L_0(t).$$

We call $L(t)$ "the osculating k-flat of X at t."

For any subset S of R^n, let $\operatorname{aff}(S)$ be the affine span of S.

DEFINITION 2. *A k-curve X is flat if $X(I)$ is included in a k-flat, i.e., if $\operatorname{aff}(X(I))$ is k-dimensional.*

THEOREM. *If all the osculating k-flats $L(t)$ of a k-curve X: $I \to R^n$ pass through a common point q, then X is flat.*

The proof of the Theorem will be given in §2–4. In §2 we make some general observations concerning k-curves. These observations imply that if the Theorem holds locally, in a neighborhood of each point $t \in I$, then it holds globally. In §3 we prove a local version of the Theorem under an extra regularity condition. In §4 we use an inductive argument to prove the local version of the Theorem without the extra condition.

The following example shows that the Theorem fails if the independence condition (b) in Definition 1 is violated even at one point.

Example. Fix k and n, $1 \le k < n$, take $I = R$, and put

$$X(t) = (x_1(t), \ldots, x_n(t)), \quad \text{where} \quad x_j(t) = t^j \quad \text{for} \quad 1 \le j \le k-1,$$

$$x_k(t) = \begin{cases} t^{k+1} & \text{for } t \ge 0 \\ 0 & \text{for } t \le 0 \end{cases}$$

$$x_{k+1}(t) = \begin{cases} 0 & \text{for } t \ge 0 \\ t^{k+1} & \text{for } t \le 0 \end{cases}$$

and $x_j(t) = 0$ for $k+2 \le j \le n$. For an example of class C^∞, replace t^{k+1} by a function $\varphi(t)$ of class C^∞ that satisfies: $\varphi^{(i)}(0) = 0$ for all $i \ge 0$, $\varphi^{(k)}(t) \ne 0$ for $t \ne 0$; e.g., the kth antiderivative of $\exp(-t^{-2})$.

2. Some preliminary observations. Let X be a k-curve.

Suppose J is an open subinterval of I, $L \subset R^n$ is a flat, and $X(J) \subset L$. Put $L_0 = L - L$. (L_0 is the translate of L through 0.) If $t \in J$, then $X^{(1)}(t)$ is the limit of scalar multiples of differences of points in $X(J)$; hence $X^{(1)}(t) \in L_0$. The same holds if t is an endpoint of J, provided $t \in I$. Repeating the same argument, we find that $X^{(i)}(t) \in L_0$ for $i = 1, 2, \ldots, k$, and for all $t \in I \cap \operatorname{cl} J$ and, therefore,

$$L(t) = X(t) + \operatorname{span}\{X^{(1)}(t), \ldots, X^{(k)}(t)\} \subset X(t) + L_0 = L.$$

It follows that $\dim L \ge k$, and if $\dim L = k$ then the last inclusion becomes an equality. Thus we have:

PROPOSITION 1. *If J is an open subinterval of I, L is a k-flat and $X(J) \subset L$, then $L = \operatorname{aff} X(J) = L(t)$ for all $t \in I \cap \operatorname{cl} J$.*

Call an open subinterval J of I *a flatness interval* (= FI) if $X(J)$ is included in a k-flat. From Proposition 1 it follows that any FI J is included in a unique maximal

FI (= MFI) $\acute{J}$. ($\acute{J}$ is the union of all FIs that include J.) Moreover, the MFIs are open and pairwise disjoint. Another consequence of Proposition 1 is

PROPOSITION 2. *If J_1 and J_2 are two disjoint FIs with a common endpoint s, then $J_1 \cup \{s\} \cup J_2$ is also a FI.*

Proof. By Proposition 1, $X(J_1 \cup \{s\} \cup J_2) \subset L(s)$. □

In particular, two distinct MFIs cannot have a common endpoint.

Denote by $\mathcal{F}$ the set of all MFIs, and put $F = \bigcup_{J \in \mathcal{F}} J$. By the previous discussion, $\mathcal{F}$ is just the set of connected components of F. In the next two sections we shall show that $F = I$. This will imply that $\mathcal{F} = \{I\}$, i.e., that I is a FI, and the theorem will be proved.

3. Proof of the theorem—part I. Recall the assumptions: $X: I \to R^n$ is a k-curve, and all the osculating k-flats $L(t)(t \in I)$ pass through a common point q. Assume, without loss of generality, that $q = 0$. (Otherwise, consider the curve $X - q$.) The meaning of the assumption $0 \in L(t)$ is: $X(t)$ is linearly dependent on $X^{(1)}(t), \ldots, X^{(k)}(t)$.

For $t \in I$ and $i = 0, 1, 2, \ldots, k$, write $X^{(i)}(t) = (x_1^{(i)}(t), \ldots, x_n^{(i)}(t))$, where $X^{(0)}(t) = X(t)$, and consider the matrix

$$\mathbf{X}(t) = \begin{bmatrix} X^{(0)}(t) \\ \vdots \\ X^{(k)}(t) \end{bmatrix} = \left(x_j^{(i)}(t): 0 \le i \le k, 1 \le j \le n\right).$$

$\mathbf{X}(t)$ is a matrix of order $(k+1) \times n$ and of rank k.

Denote by $\mathbf{X}^{\uparrow}(t)$ [$\mathbf{X}^{\downarrow}(t)$] the submatrix that consists of the first [resp., last] k rows of $\mathbf{X}(t)$. By assumption, $\mathbf{X}^{\uparrow}(t)$ is a differentiable function of t, for $t \in I$, and $\mathbf{X}^{\downarrow}(t)$ is a $k \times n$ matrix of rank k.

The next Lemma is a local version of the main theorem, with one extra condition.

LEMMA. *Suppose that $t_0 \in I$, and that the rank of $\mathbf{X}^{\uparrow}(t_0)$ is k (i.e., $X(t_0)$ is not a linear combination of $X^{(1)}(t_0), \ldots, X^{(k-1)}(t_0)$). Then t_0 lies in a flatness interval of X.*

Proof. The matrix $\mathbf{X}^{\uparrow}(t_0)$ has a nonsingular $k \times k$ submatrix, i.e., there is a set $K \subset N(= \{1, \ldots, n\})$ of size k, such that the determinant

$$\Delta(t) = \det\left(x_j^{(i)}(t): 0 \le i \le k-1,\ j \in K\right)$$

is nonzero for $t = t_0$. Since $\Delta(t)$ is continuous (even differentiable), there is an open interval J, $t_0 \in J \subset I$, such that $\Delta(t) \neq 0$ for all $t \in J$.

Consider a coordinate $l \in N \setminus K$. For $t \in J$, the columns of $\mathbf{X}(t)$ with indices in K are linearly independent, and form a basis of the column space of $\mathbf{X}(t)$. Thus the lth column of $\mathbf{X}(t)$ can be written as a linear combination of the basis columns. The coefficients of the combination can be calculated already from the submatrix $\mathbf{X}^{\uparrow}(t)$, by Cramer's Rule, as a quotient of two subdeterminants of $\mathbf{X}^{\uparrow}(t)$, where the denominator is $\Delta(t)$. Write this explicitly:

$$x_l^{(i)}(t) = \sum_{j \in K} \lambda_{lj}(t) x_j^{(i)}(t). \qquad E(i, l, t)$$

(i is the row index, $0 \le i \le k$, $l \in N \setminus K$ and $t \in J$.)

From the previous discussion it follows that the $\lambda_{lj}(t)$'s are differentiable functions of t. Differentiating equation $E(i, l, t)$, for $0 \le i \le k-1$, we obtain

$$x_l^{(i+1)}(t) = \frac{d}{dt}x_l^{(i)}(t) = \sum_{j\in K} \lambda_{lj}(t)x_j^{(i+1)}(t) + \sum_{j\in K} \dot{\lambda}_{lj}(t)x_j^{(i)}(t).$$

The first row in this equation is actually the identity $E(i+1, l, t)$. Subtracting it we get $\sum_{j\in K}\dot{\lambda}_{lj}x_j^{(i)}(t) = 0$ $(0 \le i \le k-1,\ l \in N \setminus K,\ t \in J)$.

For any fixed $l \in N \setminus K$ and $t \in J$, this is a system of k homogeneous linear equations in k unknowns $\dot{\lambda}_{lj}(t)$, and the coefficient matrix $(x_j^{(i)}(t)\colon 0 \le i \le k-1,\ j \in K)$ is nonsingular (its determinant is $\Delta(t)$). It follows that $\dot{\lambda}_{lj}(t) = 0$ for $l \in N \setminus K$, $j \in K$ and $t \in J$, so $\lambda_{lj}(t) = \lambda_{lj}$ is constant in t for $t \in J$, $l \in N \setminus K$ and $j \in K$. Thus we find that $X(J)$ is included in the linear subspace of R^n determined by the equations $E(0, l, t)$, i.e.:

$$x_l = \sum_{j\in K} \lambda_{lj}x_j, \qquad l \in N \setminus K.$$

(For $j \in K$, the coordinates x_j are free, and for $l \in N \setminus K$, x_l is a linear combination with constant coefficients of the coordinates $\{x_j\colon j \in K\}$.) □

4. Proof of the theorem—part II. Recall the notations from §2: $\mathcal{F}$ denotes the set of all MFIs, and F is their union. All we have to show is that $F = I$. By the Lemma, F contains all points $t \in I$ for which rank $\mathbf{X}^{\uparrow}(t) = k$. Define: $S = I \setminus F$. We shall complete the proof by showing that $S = \phi$.

PROPOSITION 3. *S has no isolated points.*

Proof. If s is an isolated point of S, then s is the common endpoint of two distinct components of F, i.e., of two distinct MFIs. But this is impossible, by Proposition 2. □

Now suppose $s \in S$. From the Lemma we learn that rank $\mathbf{X}^{\uparrow}(s) < k$. Since $\mathbf{X}^{\uparrow}(s)$ has k rows $X(s), X^{(1)}(s), \ldots, X^{(k-1)}(s)$, and the last $k-1$ rows are linearly independent, it follows that

$$X(s) \text{ is a linear combination of } X^{(1)}(s), \ldots, X^{(k-1)}(s). \qquad (*)$$

We shall show that $X(s)$ is, in fact, a linear combination of $X^{(1)}(s), \ldots, X^{(k-2)}(s)$ (assuming $k \ge 2$).

Take a set $K \subset N$ of size k, and consider the $k \times k$ determinant $\delta_K(t) = \det(x_j^{(i)}(t)\colon 0 \le i \le k-1,\ j \in K)$. By the Lemma, $\delta_K(t) = 0$ for all $t \in S$, in particular for $t = s$ and for a sequence $\{s_m\}_{m=1}^{\infty}$ of points in $S \setminus \{s\}$ that converge to s. (Remember: S has no isolated points!) It follows that $\dot{\delta}_K(s) = 0$. But the derivative of the $k \times k$ determinant $\delta_K(s)$ is obtained by differentiating the k rows one at a time and adding the results. Since the derivative of each row of $\mathbf{X}^{\uparrow}(s)$, except the last, is equal to the succeeding row, all that remains from the differentia-

tion is

$$0 = \delta_K(s) = \det\left(x_j^{(i)}(s)\colon i \in \{0, 1, \ldots, k-2, k\},\ j \in K\right).$$

Since this is true for all choices of K ($K \subset N$, $|K| = k$), we conclude that the matrix whose rows are $X(s)$, $X^{(1)}(s), \ldots, X^{(k-2)}(s)$, $X^{(k)}(s)$ is of rank $< k$, i.e., its rows are linearly dependent. It follows that $X(s)$ is a linear combination of the (linearly independent) rows $X^{(1)}(s), \ldots, X^{(k-2)}(s)$, $X^{(k)}(s)$. This, together with statement $(*)$ above, shows that $X(s)$ is a linear combination of $X^{(1)}(s), \ldots, X^{(k-2)}(s)$ (since the representation of $X(s)$ in terms of the linearly independent vectors $X^{(1)}(s), \ldots, X^{(k)}(s)$ is unique), and this conclusion holds, of course, for all points $s \in S$.

Repeating this argument $k-2$ more times we finally conclude that $X(s)$ depends linearly on the empty set, i.e., that $X(s) = 0$ for all $s \in S$. Differentiating once more, we find that $X^{(1)}(s) = 0$ for all $s \in S$, and therefore $S = \phi$, since by assumption $X^{(1)}(t) \neq 0$ for all $t \in I$. □

5. An open problem. Consider a continuous curve $X\colon I \to R^n$, a point $t_0 \in \operatorname{int} I$ and a k-flat $L \subset R^n$. Call L an *osculating* k-flat of X at t_0 if $X(t_0) \in L$, and

$$\operatorname{dist}(X(t_0 + h), L) = o(h^k) \quad \text{as } h \to 0. \qquad (**)$$

If $(**)$ holds only for $h < 0$ [or for $h > 0$], then we say that L is a *left* [resp., *right*] osculating k-flat of X at t_0.

Assume that X admits at each point $t \in \operatorname{int} I$ a unique left osculating k-flat $L_-(t)$, and a unique right osculating k-flat $L_+(t)$, and that $L_-(t) = L_+(t)\,(= L(t))$ for all $t \in \operatorname{int} I$. Assume also that all these osculating k-flats $L(t)$ pass through a common point. Does it follow that $X(I)$ lies in a k-flat? One might add some extra conditions, e.g., that $X(s)$ is parametrized by arc length.

It is not sufficient to assume that X admits at each point $t \in \operatorname{int} I$ a unique osculating k-flat, and that all these k-flats pass through one point. This is shown by the following example: $k = 2$, $n = 3$, $I = R$, and

$$X(t) = \begin{cases} (t, t^2, 0) & t \geq 0 \\ (t, 0, t^4) & t \leq 0 \end{cases}.$$

(At $t = 0$, the unique left osculating 2-flat is the (x_1, x_2) plane, but every plane that includes the x_1-axis is osculating on the right.)

REFERENCES

1. C. C. Hsiung, A First Course in Differential Geometry, John Wiley & Sons, 1981.
2. M. M. Lipschutz, Theory and Problems of Differential Geometry, Schaum's Outline Series, McGraw-Hill, 1969.
3. W. Klingenberg, A Course in Differential Geometry, Graduate Texts in Mathematics 51, Springer Verlag, 2nd edition, 1983.

A Counterexample on Vector Norms and the Subordinate Matrix Norms

PIETER P. N. DE GROEN
Vrije Universiteit Brussel, Department of Mathematics and Computer Science
Pleinlaan 2, B-1050 Brussels, Belgium

In an introductory course on numerical linear algebra, or in a good textbook suited for such a course (e.g., Golub and Van Loan [**1**]), one begins with some facts on vector and matrix norms. Examples of such norms are given, and it is shown that a vector norm $\|\cdot\|$ on a linear space E induces a matrix norm on the set of linear transformations from E to E, again denoted by $\|\cdot\|$,

$$\|A\| := \max_{\mathbf{x}\in E,\, \mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \qquad (A\colon E \to E).$$

This matrix norm, called the *subordinate* matrix norm, has the additional property of being *submultiplicative* with the norm of the identity equal to one,

$$\|AB\| \leqslant \|A\|\,\|B\|, \qquad \|I\| = 1 \qquad (A, B, I\colon E \to E,\ I\mathbf{x} = \mathbf{x}\ \forall \mathbf{x}).$$

Next, the Frobenius norm ($\|A\|_F^2 := \sum_{ij} a_{ij}^2$) is given as a counterexample that not every matrix norm is subordinate to some vector norm. However, the Frobenius norm of the identity is equal to the square root of the dimension of the underlying vector space E, so it does not falsify the converse question: *Is every submultiplicative matrix norm, in which the norm of the identity is equal to one, subordinate to a vector norm*? If you pose this question to a colleague, he or she immediately answers "no" (unless they consider the $n \times n$ matrix A as a transformation of the n^2-dimensional space of $n \times n$ matrices by matrix multiplication). If you continue and ask for a counterexample, they are unable to supply either such an example or a reference in which it could be found. Here is a construction of a simple counterexample.

Let us assume that $\|\cdot\|_\lambda$ is the following matrix norm on $\mathbb{R}^{2\times 2}$, depending on a positive parameter λ,

$$\|A\|_\lambda := \max\{|a_{11}| + \lambda|a_{12}|, |a_{22}| + \lambda|a_{21}|\}, \qquad A := \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

If $\lambda = 1$, this is the matrix norm subordinate to the vector *sup*-norm. We shall show, that it satisfies the requirements mentioned above and that it cannot be subordinate to a vector norm for any $\lambda > 1$.

The strict positivity, the triangle inequality, and the homogeneity of degree 1 are trivially satisfied by this matrix function, implying that it is a norm. Moreover, the norm of the identity is equal to one. For the submultiplicativity we consider matrices A and B,

$$A := \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad B := \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

For simplicity we assume

$$\|A\|_\lambda = |a_{11}| + \lambda|a_{12}| \geqslant |a_{22}| + \lambda|a_{21}|,$$
$$\|B\|_\lambda = |b_{11}| + \lambda|b_{12}| \geqslant |b_{22}| + \lambda|b_{21}|.$$

For the absolute row sum of the first row of the product AB we then find:

$$|a_{11}b_{11}+a_{12}b_{21}|+\lambda|a_{11}b_{12}+a_{12}b_{22}| \leqslant |a_{11}|(|b_{11}| + \lambda|b_{12}|) + \lambda|a_{12}|(|b_{22}| + \lambda|b_{21}|)$$
$$\leqslant \|A\|_\lambda \|B\|_\lambda,$$

provided $\lambda \geqslant 1$. Since the same holds for the sum of the second row, this implies that the proposed norm is submultiplicative for any $\lambda \geqslant 1$.

Let us now assume that there exists a vector norm, to be denoted by $\|\cdot\|$, such that $\|\cdot\|_\lambda$ is subordinate to this vector norm. We first show that this norm must be *monotone*. Let $e_1 := (1,0)^T$ and $e_2 := (0,1)^T$ be the "unit" base vectors in $\mathbb{R}^2$ and let A and B be the matrices

$$A := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad B := \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

For all vectors $\mathbf{z} \in \mathbb{R}^2$, $\mathbf{z} = (z_1, z_2)^T$ we have

$$1 = \|A\|_\lambda \geqslant \frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} = |z_1|\frac{\|\mathbf{e}_1\|}{\|\mathbf{z}\|}, \quad \Rightarrow \|\mathbf{z}\| \geqslant |\mathbf{z}_1|\,\|\mathbf{e}_1\|.$$

Using B we likewise can show: $\|\mathbf{z}\| \geqslant |\mathbf{z}_2|\,\|\mathbf{e}_2\|$, for all $\mathbf{z} \in \mathbb{R}^2$.

Now we can derive a contradiction in the case $\lambda > 1$. Let C and D be the matrices

$$C := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad D := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Since the matrix norm $\|\cdot\|_\lambda$ is subordinate to the vector norm $\|\cdot\|$, a nonnull vector $\mathbf{x}$ exists, such that

$$\lambda = \|C\|_\lambda = \frac{\|C\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{|x_2|}{\|\mathbf{x}\|}\|\mathbf{e}_1\|.$$

Hence the monotonicity of the norm implies

$$|x_2|\,\|\mathbf{e}_1\| = \lambda\|\mathbf{x}\| \geqslant \lambda|x_2|\,\|\mathbf{e}_2\|, \quad \Rightarrow \|\mathbf{e}_1\| \geqslant \lambda\|\mathbf{e}_2\|.$$

Likewise we show with the aid of the matrix D that $\|\mathbf{e}_2\| \geqslant \lambda\|\mathbf{e}_1\|$, which is impossible for any $\lambda > 1$! Clearly the matrix norm is not subordinate to any vector norm.

Finally we find that this example fits nicely in a more general theory and clarifies Stoer's [2] characterization of the subordinate matrix norms as the *minimal* (*sub*)*multiplicative* ones. A submultiplicative matrix norm $\nu(\cdot)$ is called *minimal* if any other submultiplicative norm $\mu(\cdot)$ satisfying $\mu(A) \leqslant \nu(A)$ for all matrices A is equal to ν. Obviously, in the class of $\|\cdot\|_\lambda$-norms the *sup*-norm is the minimal element.

REFERENCES

1. G. H. Golub and Ch.F. Van Loan, Matrix Computations, Johns Hopkins Univ. Press, Baltimore, 1983.
2. J. Stoer, On the characterization of least upper bound norms in matrix space, *Num. Math.*, 6 (1964) 302–314.

Ambiguous Numbers Are Dense

MARKO PETKOVŠEK
Department of Mathematics, University of Ljubljana, Jadranska 19, 61111 Ljubljana, Slovenia, Yugoslavia

1. Introduction. It is well known that in "classic" radix number systems with integer base $b \geqslant 2$ and digits $0, 1, \ldots, b-1$ every positive b-ary fraction has two distinct representations: for example,

$$1 \cdot 000\ldots_{(10)} = 0 \cdot 999\ldots_{(10)} = 1.$$

Are there number systems in which every real number has a unique representation? The purpose of this note is to show that for any real positional number system $\mathscr{S}$,

1) if $\mathscr{S}$ can represent all real numbers, then the set of numbers representable in $\mathscr{S}$ in more than one way is dense in $\mathbb{R}$, and
2) if representation in $\mathscr{S}$ is unique, then the set of numbers not representable in $\mathscr{S}$ is uncountable and dense in $\mathbb{R}$.

The proof of these results is an instructive exercise in elementary point-set topology. As often in discrete mathematics, the relevant topology is that of a Stone space (= compact, totally disconnected Hausdorff space).

Note that the situation is quite different with representation of integers, since in this case, there are many number systems in which every integer is uniquely representable (see Example 2.4). For related work concerning the ambiguity problem for representation of integers, see Maurer, Salomaa and Wood [**9**], Culik II and Salomaa [**2**], and Honkala [**3**, **4**, **5**].

2. Positional number systems

2.1 DEFINITION. *By a real positional number system we mean a two-way sequence of sets of positional values* $\langle P_k \rangle_{k \in \mathbb{Z}}$, *where*

(i) *for every* $k \in \mathbb{Z}$, P_k *is a finite set of reals containing* 0,

(ii) *the series*

$$\sum_{k<0} \max_{p_k \in P_k} |p_k|$$

is convergent.

2.2 *Example.* In our conventional decimal system, $P_k = \{d \cdot 10^k | d = 0, 1, \ldots, 9\}$. Note that what we call positional values here are not digits but rather the numerical values represented by them. For instance, the (decimal) number 32 has digits 3 and 2, but it has positional values 30 and 2.

2.3. If $\mathscr{S} = \langle P_k \rangle_{k \in \mathbb{Z}}$ is a real positional number system, let

$$R_n := \prod_{k>n} \{0\} \times \prod_{k \leqslant n} P_k, \quad \text{for } n \in \mathbb{Z},$$

and

$$R := \bigcup_{n \in \mathbb{Z}} R_n.$$

Thus R is the set of all two-way sequences of positional values in which all positions from some place on (in direction of increasing subscripts) are filled with zeros.

Define the *valuation function* v: $R \to \mathbb{R}$ by setting

$$v(p) := \sum_{k \leqslant n} p_k,$$

where $p = \cdots 00 p_n p_{n-1} p_{n-2} \cdots$ is a sequence from R_n. By Definition 2.1(ii), the series on the right-hand side is absolutely convergent. The number $v(p)$ is the *value* of p, and p is a *representation* of $v(p)$.

A representation is *finite* if $p_k = 0$ for all $k < n$, for some integer n, and *integral* if $p_k = 0$ for all $k < 0$.

A number is *representable* in $\mathscr{S}$ if it belongs to $v(R)$, and *ambiguous* in $\mathscr{S}$ if it has more than one representation in $\mathscr{S}$.

The system $\mathscr{S}$ is *complete over* $\mathbb{R}$ if $v(R) = \mathbb{R}$, and *unambiguous* if v is one-to-one.

2.4 *Example.* Let D be a finite set of reals containing 0, and $|b| > 1$. Let $\mathscr{S}(b, D)$ denote the radix system with base b and digit set D, that is, the positional number system with $P_k = b^k \cdot D$, for $k \in \mathbb{Z}$. Of prime interest are of course *integer radix systems* where the base and the digits are integers. In Matula [**7**, **8**] a digit set $D \subset \mathbb{Z}$ is called *basic for* $b \in \mathbb{Z}$ if every integer has a unique integral representation in $\mathscr{S}(b, D)$. The following are some results about basic digit sets from [**7**] and [**8**]:

1. If D is basic for b then $\mathscr{S}(b, D)$ is complete over $\mathbb{R}$.
2. If D is basic for b then D is a complete residue system mod $|b|$.
3. If D consists of $|b|$ contiguous integers which include 0, and either D contains positive and negative values, or b is negative, or both, then D is basic for b.
4. There are, however, many noncontiguous basic digit sets. In fact, if $|b| \geqslant 3$ then there are infinitely many basic digit sets for b; for instance, the sets $\{-1, 0, 3^n - 2\}$ are basic for $b = 3$, for all positive integers n.
5. If D is basic for b then the set of numbers ambiguous in $\mathscr{S}(b, D)$ is infinite, and may even be uncountable if the digits are not contiguous. Actually, the construction given in [**8**, Lemma 11] shows that this set is dense in $\mathbb{R}$.

Thus the balanced ternary system with $b = 3$ and $D = \{-1, 0, 1\}$, the "negaternary" system with $b = -3$ and $D = \{0, 1, 2\}$, and the system with $b = 3$ and $D = \{-1, 0, 7\}$, are all complete over $\mathbb{R}$. Here are some examples of ambiguity in these systems:

$$0 \cdot 111\ldots_{(3)} = 1 \cdot (-1)(-1)(-1)\ldots_{(3)} = 1/2$$

$$1 \cdot 2020\ldots_{(-3)} = 0 \cdot 0202\ldots_{(-3)} = 1/4$$

$$0 \cdot 777\ldots_{(3)} = (-1)7 \cdot (-1)(-1)(-1)\ldots_{(3)} = 7/2.$$

2.5 *Example.* As an example of a so-called *mixed-radix system*, consider the *factorial system* where the role of the powers is played by factorials and their reciprocals $(\ldots 3!, 2!, 1!; 1/2!, 1/3!, \ldots)$. The digits allowed with $n!$ or with $1/(n+1)!$ are $0, 1, \ldots, n$, so that

$$P_k = \left\{ d \cdot (|k| + 1)!^{\operatorname{sgn} k} \middle| d = 0, 1, \ldots, m_k \right\},$$

where

$$m_k = \begin{cases} k+1, & \text{if } k \geqslant 0; \\ -k, & \text{if } k < 0. \end{cases}$$

This system and its generalizations were treated by Cantor in [**1**]. It is not difficult to see that every nonnegative real number is representable in the factorial system, that every nonnegative rational number has a finite representation in it, and that every positive rational is ambiguous in it. For example,

$$1 \cdot 000\ldots_{(f)} = 0 \cdot 1234\ldots_{(f)} = 1.$$

For further examples of real and complex positional number systems, and for a history of number representations, see Knuth [**6**, pp. 162–180].

3. Ambiguity and completeness of positional systems. Let $\mathcal{S}$ be a real positional number system. For each $n \in \mathbb{Z}$, let $v_n\colon R_n \to v(R_n)$ be the restriction of v to R_n.

3.1 LEMMA. *If for some $n \in \mathbb{Z}$ the mapping v_n is one-to-one, then the set of representable numbers $v(R)$ is meager.*

Proof. Endow each set of positional values P_k, as well as the singleton $\{0\}$, with the discrete topology. Being a product of finite discrete spaces, each R_n becomes a compact totally disconnected Hausdorff space. Using Definition 2.1(ii), one sees easily that the functions v_n are continuous. As they map a compact space into a Hausdorff space, they are closed as well. Hence each $v(R_n)$ is a closed set of reals.

Assume that v_n is one-to-one. Because it is continuous and closed, it is a homeomorphism of R_n onto $v(R_n)$. Therefore, $v(R_n)$ is totally disconnected. As a set of reals, it has empty interior.

Let

$$F_n := \Big(\prod_{k>n} P_k \times \prod_{k\leqslant n} \{0\}\Big) \cap R.$$

By Definition 2.1(i), $F_n \cap R_k$ is finite, for every $k \in \mathbb{Z}$. This implies that F_n is countable. Obviously, every representable number can be written as a sum of a number from $v(F_n)$ and a number from $v(R_n)$. Therefore,

$$v(R) = v(F_n) + v(R_n) = \bigcup_{x\in v(F_n)} (x + v(R_n)).$$

This is a countable union of translates of $v(R_n)$, which is a closed set with empty interior. It follows that $v(R)$ is a meager set.

3.2 THEOREM. (i) *If $\mathcal{S}$ is complete over $\mathbb{R}$, then the set of numbers ambiguous in $\mathcal{S}$ is dense in $\mathbb{R}$.*

(ii) *If $\mathcal{S}$ is unambiguous, then the set of numbers not representable in $\mathcal{S}$ is uncountable and dense in $\mathbb{R}$.*

Proof. Recall that by Baire's Category Theorem, $\mathbb{R}$ is not a meager set.

(i) If $\mathcal{S}$ is complete then $v(R) = \mathbb{R}$ and is therefore not meager. Hence it follows from Lemma 3.1 that no v_n is one-to-one. Choose $x \in \mathbb{R}$ and $\varepsilon > 0$. As $\mathcal{S}$ is complete, $x = v(p)$ for some $p = \cdots 00p_n p_{n-1} \cdots$ By continuity of v_n, there is an $m \in \mathbb{Z}$ such that $y + v(R_m) \subseteq (x-\varepsilon, x+\varepsilon)$ where $y := v(\cdots 00p_n p_{n-1} \cdots p_{m+1}00\cdots)$. Since v_m is not one-to-one, R_m con-

tains two distinct representations of some $z \in v(R_m)$. But then $(y + z) \in (x - \varepsilon, x + \varepsilon)$ is also ambiguous. Thus we have shown that every interval of positive length contains ambiguous numbers, which means that they are dense in $\mathbb{R}$.

(ii) If $\mathscr{S}$ is unambiguous, then each v_n is one-to-one. By Lemma 3.1, it follows that $v(R)$ is a meager set of reals. Hence its complement $\mathbb{R} \setminus v(R)$ is an uncountable dense set.

3.3 *Remark.* It is possible to modify these results for systems with nonnegative positional values (such as the "classic" radix systems, or the factorial system), by everywhere replacing $\mathbb{R}$ with $\mathbb{R}^+ \cup \{0\}$.

Acknowledgement. The author wishes to express his thanks to the referees for their valuable suggestions. In particular, the present title of the note is due to one of them.

REFERENCES

1. G. Cantor, Über die einfachen Zahlensysteme, *Zeitschrift für Mathematik und Physik*, 14 (1869) 121–128.
2. K. Culik II and A. Salomaa, Ambiguity and decision problems concerning number systems, *Information and Control*, 56 (1983) 139–153.
3. J. Honkala, Bases and ambiguity of number systems, *Theoret. Comput. Sci.*, 31 (1984) 61–71.
4. _____, A decision method for the recognizability of sets defined by number systems, *RAIRO Informatique théorique et Applications*, 20 (1986) 395–403.
5. _____, On number systems with negative digits, *Ann. Acad. Sci. Fenn. Ser. AI Math.*, 14 (1989) 149–156.
6. D. E. Knuth, The Art of Computer Programming, Vol. 2: Seminumerical Algorithms, Addison-Wesley, Reading, Mass., 1969.
7. D. W. Matula, Radix arithmetic: Digital algorithms for computer architecture, in Applied Computation Theory: Analysis, Design, Modeling, R. T. Yeh, Ed., Prentice-Hall, Englewood Cliffs, N.J., 1976, pp. 374–448.
8. _____, Basic digit sets for radix representation, *J. ACM*, 29 (1982) 1131–1143.
9. H. A. Maurer, A. Salomaa and D. Wood, L codes and number systems, *Theoret. Comput. Sci.*, 22 (1983) 331–346.

The Diophantine Equation $x^2 + y^m = z^{2n}$, Addendum

DAVID W. BOYD
Department of Mathematics, University of British Columbia, Vancouver, B.C., Canada V6T 1Y4

Professor Andrzej Schinzel has kindly informed me that the main result of [**1**], that the equation of the title has infinitely many primitive solutions if $(m, n) = 1$, follows from Corollary 2 of his paper [**2**, p. 49]. A special case of his general result is that the equation in question is birationally equivalent to $X^2 + Y^{(m,2)} = Z^2$. Since the trivial equation $X^2 + Y = Z^2$ and the Pythagorean equation each have infinitely many primitive solutions, the result follows immediately.

The idea is to multiply the equation $X^2 + Y^{(m,2)} = Z^2$ by $Y^{2a}Z^{2b}$ where a and b are chosen so that the three terms become the desired powers. This is possible by the Chinese remainder theorem. For example, to obtain $x^2 + y^3 = z^4$, multiply

$X^2 + Y = Z^2$ by Y^8Z^6 and then define $(x, y, z) = (XY^4Z^3, Y^3Z^2, Y^2Z^2)$. For example, $1^2 + 3 = 2^2$ gives the solution $(2^3 3^4, 2^2 3^3, 2^2 3^2)$ of $x^2 + y^3 = z^4$.

Conversely, a solution of $x^2 + y^3 = z^4$ yields the solution $(X, Y, Z) = (x, y^3, z^2)$ of $X^2 + Y = Z^2$. Applying the process of the previous paragraph to (X, Y, Z) leads to a solution (x', y', z') of $x^2 + y^3 = z^4$ which is in the same class as the original (x, y, z), showing that this method yields all (classes of) solutions of the equation of interest.

The approaches of [1] and [2] are quite different but both are suitable for inclusion in an elementary number theory course. One obvious advantage to the approach used in [2] is that the number of parameters needed to represent all solutions does not increase with m.

REFERENCES

1. David W. Boyd, The diophantine equation $x^2 + y^m = z^{2n}$, *Amer. Math. Monthly*, 95 (1988) 544–547.
2. A. Schinzel, O równaniu diofantycznym $\Sigma A_k x_k^{\theta_k} = 0$, *Prace Mat.*, 4 (1960) 45–51. (Polish with Russian and French abstracts.)

Errata, addenda, etc.

1. Dr. Arthur T. Benjamin informs us that the main result of his note The bisection method: which root?, this MONTHLY 94 (1987) 861–863, had appeared previously in a note by George Corliss, Which root does the bisection algorithm find?, *SIAM Rev.*, April, 1977.
2. Jingcheng Wong, of the University of North Florida, has noted an error in On the mean value theorem for integrals by B. Jacobson, which appeared in the May, 1982 issue, pp. 300–301. Theorem 2 of that article is false. The limit $c \to a$ which appears in several places on p. 301 should read $x \to a$.
3. Dr. K. D. Magill advises us that since the publication of his article (Universal topological spaces 95 (1988) 942–946) he has learned that theorem 1 had previously appeared in On a certain class of complete regularity, by J. Ferrer and V. Gregori, *Acta. Math. Hung.*, 47 (1986) 179–180.
4. Dr. John Beebee writes that he has found an error in his note Examples of infinite incongruent exact covers, 95 (1988) 121–123. The set of arithmetic progressions $\{(2^i : 2^{i-1}) | i \geq 1\}$ does not cover 0. The error can be corrected by changing the offsets, and the conclusion is unchanged.
5. I am advised by Mr. Hans-J. Kleiner, of Spaichingen, F.R.G., that there is an error in the factored form of the chromatic polynomial of the icosahedron that appeared in D. H. Lehmer's article Coloring the platonic solids (93, 288–292). It should read as '$\ldots + 6999t^4 - 19698t^3 + \ldots$' instead of '$\ldots + 6999t^4 - 19688t^3 + \ldots$'.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

Singular Values and the Spectral Theorem

K. HOECHSMANN
Department of Mathematics, University of B.C., Vancouver V6T 1Y4, Canada

The action of an arbitrary real or complex $m \times n$-matrix is not at all arbitrary: it takes a suitable orthonormal basis of n-space into some orthogonal set in m-space; consequently it maps the unit n-ball onto a (possibly lower-dimensional) ellipsoid. This is the geometric view of the Singular Value Theorem, which is not only one of the nicest matrix theorems to state and to visualize but also one of the easiest to prove and to apply. In any introductory course on matrices it deserves a place near the center. Theorem 1 below is a pure matrix version of this result. One of its main consequences, Theorem 2 below, leads straight to the orthogonal diagonalization of certain matrices, i.e., the Spectral Theorem.

It is customary to prove Theorem 2 independently, applying either the Fundamental Theorem of Algebra to a characteristic polynomial or Lagrange Multipliers to a quadratic form, and even to deduce Theorem 1 from *it*. The analytic equipment needed for the following proof is more modest: it suffices to know that a continuous real-valued function on any compact set has a maximum. For the sake of simplicity the argument will first be given for *real* matrices and then (trivially) extended to complex ones. To prevent any suspicion that extraneous subtleties are tacitly used, the prefix "eigen" will be avoided.

THEOREM 1. *Let $A \neq 0$ be an $m \times n$ real matrix. Then there exist orthogonal matrices M and N such that*

$$MAN = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \quad \text{where} \quad D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_r \end{bmatrix} \quad \text{with} \quad d_i \geq d_{i+1} > 0.$$

Proof. Let $O(k)$ be the set of all $k \times k$ orthogonal matrices. For $M \in O(m)$ and $N \in O(n)$, let $\alpha(M, N)$ stand for the entry in the first row and first column of MAN. Clearly, α is a continuous function of the pair (M, N) and, therefore, attains a maximal value $d_1 > 0$ on the set $O(m) \times O(n)$, which is closed and bounded in a suitable Euclidean space. Say $d_1 = \alpha(M_1, N_1)$. Then

$$M_1 A N_1 = \begin{bmatrix} d_1 & X \\ Y & A' \end{bmatrix},$$

where A' is an $(m-1) \times (n-1)$ matrix. It turns out that X and Y are actually *zero* rows and columns, respectively. Indeed, if X were nontrivial, the first row ρ_1 of $M_1 A N_1$ would have length $d > d_1$. Then one could multiply on the right by the reflection H which takes ρ_1 to $[d, 0, \ldots, 0]$ and create a value $\alpha(M, N) = d > d_1$. For similar reasons, involving columns and left multiplication, it follows that $Y = 0$. Finally, none of the entries of A' can exceed d_1 in absolute value (because each of

them could be permuted to the upper left), and this is true for all the possible forms of A'. An obvious induction now finishes the proof.

The positive numbers $d_1 \geqslant \cdots \geqslant d_r$ are known as the *singular values* of A.

Remarks. The $n \times n$ matrix $A^{\#} = A^T A$ has a very simple effect on the columns $v_1, \ldots, v_n$ of N; in fact $A^{\#} v_i = \mu_i v_i$, where $\mu_i = d_i^2$ for $i \leq r$, and $= 0$ beyond. In matrix terms, this says that $A^{\#} N = N\Delta$, where Δ is a diagonal matrix with diagonal entries μ_i as described. It is proved by putting $\Delta = (MAN)^T (MAN)$, which is just $N^T A^{\#} N$.

To prove *uniqueness* of the singular values d_i it clearly suffices to characterize the μ_i as being the only numbers such that $(A^{\#} - \mu I)v = 0$ for some $v \neq 0$. Indeed, setting $v = \Sigma a_i v_i$, one gets $(A^{\#} - \mu I)v = \Sigma a_i(\mu_i - \mu)v_i$, which is never 0, unless μ is one of the μ_i. More geometrically, the d_i can also be retrieved from the image under A of the appropriate unit sphere.

If $m = n$, the theorem is often written as $A = SR$, with S positive semidefinite and $R = M^T N^T$ orthogonal. It is then called the polar decomposition.

THEOREM 2. *Every real symmetric $n \times n$ matrix A has an invariant one-dimensional subspace.*

Proof. Let $v \neq 0$ be one of the columns of N, so that $(A^{\#} - \mu I)v = 0$, as in the remarks above. The symmetric nature of A makes $A^{\#} = A^T A = A^2$ and

$$0 = (A^{\#} - \mu I)v = (A + \lambda I)(A - \lambda I)v,$$

where $\lambda^2 = \mu$. If $(A - \lambda I)v = w \neq 0$, then w generates such a line; otherwise v does.

Remarks. For symmetric A it is trivial to show that the orthocomplement of any invariant subspace is itself invariant. Hence, by induction, Theorem 2 provides a set of n mutually orthogonal invariant lines, thus proving the spectral theorem for symmetric matrices. Moreover, if B is symmetric and commutes with A, it defines a symmetric operator on the non-zero kernel of $A - \lambda I$. Hence, the two matrices have a *common* invariant line, and again, by induction, a complete orthogonal set of such lines.

If the use of linear transformations is didactically impracticable, the spectral theorem can be stated in terms of an orthogonal U making $U^T A U$ diagonal. Then the induction will hinge on partial diagonalizations $V^T A V$, where the columns of V form an orthonormal basis of an invariant subspace and its orthocomplement.

Everything said after the first sentence of Theorem 1 and before the present one applies *verbatim* to complex matrices, if one changes "orthogonal" to "unitary," "symmetric" to "hermitian," and replaces the transpose A^T by its complex conjugate A^* (adjoint of A). Writing every complex matrix as $C = A + iB$, with A and B hermitian, one can clearly extend the spectral theorem to all C for which A and B commute, i.e., where C commutes with its adjoint $A - iB$. These are, of course, the *normal* matrices.

A Remark on the Shortest Confidence Interval of a Normal Mean

ROBERT BARTOSZYŃSKI
Department of Statistics, Ohio State University, Columbus, OH 43210

WAI CHAN
Digital Equipment Corporation, AET1-2 / 7, Andover, MA 01810

1. Introduction. In estimation theory, the shortest 95% confidence interval for the mean of a normal distribution of known variance σ^2 is given by

$$\bar{X} \pm 1.96\sigma/\sqrt{n} \tag{1}$$

(see, for example, Lehmann [1]).

Dr. Jones, a recent graduate from the *Department of Dubious Statistics*, found an even shorter confidence interval. First he computed the interval given in (1). He then pretended that σ is unknown and computed the interval

$$\bar{X} \pm tS/\sqrt{n-1}, \tag{2}$$

where $nS^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2$ and $t = t_{.975}^{(n-1)}$ is the 97.5th percentile of a student's t-distribution with $n-1$ degrees of freedom. Now he chooses whichever of intervals (1) and (2) that happened to be shorter.

Obviously, the length of Dr. Jones' interval is less than or equal to the length of interval (1). On the other hand, both (1) and (2) have coverage probability of .95, so Dr. Jones claims that the same must be true for his interval.

Has Dr. Jones made a significant statistical discovery?

This question, plus others of a more technical nature, was on the so-called Inference Exam in the Department of Statistics of Ohio State University in 1987.

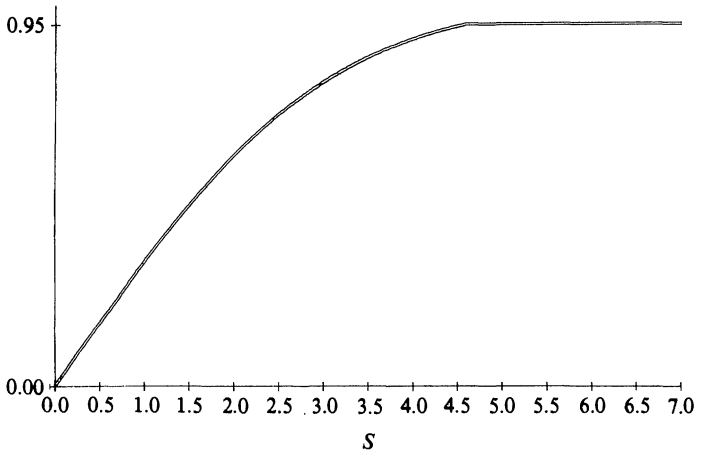


FIG. 1. P (Coverage$|S = s$)

Variance = 25
n = 21

We cannot disclose the details of students' performance. However, the question here is not entirely without curiosity value. Even experienced statisticians are known to have been baffled by Dr. Jones' reasoning momentarily. The important point here is to note that even though the probability $P(|\bar{X} - \mu| \leqslant tS/\sqrt{n-1})$ is .95, the conditional probability $P(|\bar{X} - \mu| \leqslant tS/\sqrt{n-1} \mid S = s)$ is not constant in s. When $ts/\sqrt{n-1} < 1.96\sigma/\sqrt{n}$, the conditional probability $P(|\bar{X} - \mu| \leqslant tS/\sqrt{n-1} \mid S = s)$ is less than .95. In Figure 1, we graph the conditional coverage probability of Dr. Jones' interval as a function of s for $n = 21$ and $\sigma^2 = 25$. Thus while the expected length of Dr. Jones' interval is less than $2 \times 1.96\sigma/\sqrt{n}$ the coverage probability is less than .95. We shall briefly describe the computations of the expected length and coverage probability of Dr. Jones' interval.

2. Expected length. The length of the newly defined confidence interval is a function of S and is therefore a random variable. To calculate this expected length, let f_{n-1} denote the density function of a chi-square random variable with $n - 1$ degrees of freedom. Then,

$$
\begin{aligned}
&E(|\text{Jones' interval}|) \\
&= \int_{\{t^2s^2/(n-1) \geqslant (1.96\sigma)^2/n\}} 2 \times 1.96\sigma/\sqrt{n} \times f_{n-1}\left(\frac{ns^2}{\sigma^2}\right) d\left(\frac{ns^2}{\sigma^2}\right) \\
&\quad + \int_{\{t^2s^2/(n-1) < (1.96\sigma)^2/n\}} 2 \times ts/\sqrt{n-1} \times f_{n-1}\left(\frac{ns^2}{\sigma^2}\right) d\left(\frac{ns^2}{\sigma^2}\right) \qquad (3) \\
&= 2 \times 1.96\sigma/\sqrt{n} \times P\left(nS^2/\sigma^2 \geqslant (n-1)(1.96/t)^2\right) \\
&\quad + 2t/\sqrt{n-1} \int_{\{(ns/\sigma)^2 < (n-1)(1.96/t)^2\}} s \times f_{n-1}\left(\frac{ns^2}{\sigma^2}\right) d\left(\frac{ns^2}{\sigma^2}\right). \qquad (4)
\end{aligned}
$$

The first term of (4) is

$$
\left(2 \times 1.96\sigma/\sqrt{n}\right) P\left(\chi^2_{n-1} \geqslant (n-1)(1.96/t)^2\right). \qquad (5)
$$

By a change of variable, the second term of (4) becomes

$$
\begin{aligned}
&\frac{2t}{\sqrt{n(n-1)}} \int_0^{(n-1)(1.96/t)^2} \sigma\sqrt{y} f_{n-1}(y)\, dy \\
&= 2t\sigma \sqrt{\frac{2}{n(n-1)}} \times \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \times P\left(\chi^2_n \leqslant (n-1)(1.96/t)^2\right). \qquad (6)
\end{aligned}
$$

Thus we need only a chi-square table to compute the expected length of the interval. For example if $n = 21$ and $\sigma^2 = 25$, then using (5) and (6) we obtain,

$$
E(|\text{Jones' interval}|) = 2.61 + 1.48 = 4.09,
$$

whereas the length of the standard confidence interval of (1) is given in this case as 4.28.

3. Coverage Probability. The coverage probability is

$$\begin{aligned} &P(\text{COVERAGE}) \\ &\quad = P\left(|\bar{X} - \mu| \leqslant 1.96\sigma/\sqrt{n}\,;\, 1.96\sigma/\sqrt{n} \leqslant tS/\sqrt{n-1}\right) \\ &\qquad + P\left(|\bar{X} - \mu| \leqslant tS/\sqrt{n-1}\,;\, 1.96\sigma/\sqrt{n} > tS/\sqrt{n-1}\right). \end{aligned}$$

Now by conditioning on the random variable S and using the fact that $\bar{X}$ and S are independent, we obtain

$$\begin{aligned} &P(\text{COVERAGE}) \\ &= \int_{\{(1.96\sigma)^2/n \leqslant t^2 s^2/(n-1)\}} P\left(|\bar{X} - \mu| \leqslant 1.96\sigma/\sqrt{n} \,|\, S = s\right) f_{n-1}\left(\frac{ns^2}{\sigma^2}\right) d\left(\frac{ns^2}{\sigma^2}\right) \\ &\quad + \int_{\{(1.96\sigma)^2/n > t^2 s^2/(n-1)\}} P\left(|\bar{X} - \mu| \leqslant ts/\sqrt{n-1} \,|\, S = s\right) f_{n-1}\left(\frac{ns^2}{\sigma^2}\right) d\left(\frac{ns^2}{\sigma^2}\right) \qquad (7) \\ &= .95P\left(nS^2/\sigma^2 \geqslant (n-1)(1.96/t)^2\right) \\ &\quad + \int_{\{ns^2/\sigma^2 < (n-1)(1.96/t)^2\}} P\left(|\bar{X} - \mu| \leqslant ts/\sqrt{n-1}\right) f_{n-1}\left(\frac{ns^2}{\sigma^2}\right) d\left(\frac{ns^2}{\sigma^2}\right). \qquad (8) \end{aligned}$$

The first term of (8) is $.95P(\chi^2_{n-1} \geqslant d)$ where $d = (n-1)(1.96/t)^2$. The second term of (8), by a change of variable, becomes

$$\begin{aligned} &\int_{y=0}^{d} P\left(|\bar{X} - \mu| \leqslant t\sigma\sqrt{y/(n^2 - n)}\right) f_{n-1}(y)\, dy \\ &\qquad = \int_{y=0}^{d} \left(2\Phi\left(t\sqrt{y/(n-1)}\right) - 1\right) f_{n-1}(y)\, dy, \end{aligned}$$

where Φ is the CDF of a standard normal distribution. Thus (8) can be rewritten as

$$\begin{aligned} &P(\text{COVERAGE}) \\ &\quad = .95 - 1.95P\left(\chi^2_{n-1} \leqslant d\right) + 2\int_{y=0}^{d} \Phi\left(t\sqrt{y/(n-1)}\right) f_{n-1}(y)\, dy, \qquad (9) \end{aligned}$$

which can be evaluated numerically. For $n = 21$ and $\sigma^2 = 25$, the coverage probability is .9358. Note that whenever $y < d = (n-1)(1.96/t)^2$, $\Phi(t\sqrt{y/(n-1)}) < \Phi(1.96) = .975$. From (9), we obtain,

$$\begin{aligned} &P(\text{COVERAGE}) \\ &\quad < .95 - 1.95P\left(\chi^2_{n-1} \leqslant d\right) + 2\int_0^d .975 f_{n-1}(y)\, dy \\ &\quad = .95 - 1.95P\left(\chi^2_{n-1} \leqslant d\right) + 2(.975)P\left(\chi^2_{n-1} \leqslant d\right) \\ &\quad = .95, \qquad (10) \end{aligned}$$

and so the coverage probability of Dr. Jones' interval is less than .95 as asserted.

REFERENCE

1. E. L. Lehmann, *Testing Statistical Hypotheses*, Wiley, New York, 1959.

For Your Eyes Only

GARY A. MARTIN
Department of Mathematics, Southeastern Massachusetts University, North Dartmouth, MA 02747

I was startled to learn from James C. Reber's article, "Another Required Reading Program for Mathematics Majors" [1], that sections from Guillen's *Bridges to Infinity* [2] are occasionally required reading for students taking Calculus III at Indiana University of Pennsylvania. I see no possible reason for encouraging potential mathematics majors to read this book unless it is to point out the numerous substantive errors it contains. To be fair, some of the sections seem to be free of misinformation; on the other hand, they also appear to be free of mathematical content. If it is these which are assigned, the danger is that the interested student will read the rest on her or his own initiative, and enter advanced courses believing some of the following assertions:

(a) "On the rational number line, adjacent numbers are infinitely close together. But with the real number line adjacent numbers are *closer* than infinitely close together." (page 35)

(b) "It $[2^{\aleph_0}]$ is the first stepping stone beyond infinity, the first transfinite number, which Cantor named $\aleph_1$. A set with $\aleph_1$ elements, in turn, has precisely $2^{\aleph_1}$ conceivable subsets. This is the second stepping stone, the second transfinite number $\aleph_2$, and so forth."[1] (page 46)

(c) "If it were possible to write out explicitly the zeroth transfinite number, $\aleph_0$, it would be a 1 followed by an infinity of zeros. The transfinite number $\aleph_1$ would be a 1 followed by more than an infinity of zeros." (page 47)

(d) "With such a set [the limit of the $\aleph_n$s, which he calls $\aleph_\infty$], he [Cantor] could recycle his [diagonal] argument to define an entirely new sequence of larger-than-$\aleph_\infty$, or trans-transfinite, sets. One might name them after the second Hebrew letter, ב (beth), so that the new sequence would be $\beth_0, \beth_1, \beth_2$, and so on."[2] (page 47)

(e) "There are mathematicians today who do speak of an "absolute infinite." They denote it with the last letter of the Greek alphabet, omega (ω), and impute it to be the largest conceivable infinity."[3] (page 47)

(f) "To this day we still don't know exactly how many irrationals there are, although it has been established that the total number cannot be more than $\aleph_1$. Cantor himself guessed that the total number of irrationals is exactly $\aleph_1$, mainly since $\aleph_1$ is the next largest infinity after $\aleph_0$ defined by set theory. His guess came to be known as the Continuum Hypothesis. But there is still the uneliminated possibility that the number of irrationals actually lies somewhere between $\aleph_0$ and $\aleph_1$."[4] (page 50)

[1] The author is incorporating the Generalized Continuum Hypothesis into his notation!

[2] We know these as $\aleph_\omega, \aleph_{\omega+1}, \aleph_{\omega+2}, \ldots$ —$\beth_n$ is obtained by iterating the power set operator n times, beginning with $\aleph_0$.

[3] We use ω to denote the first infinite ordinal—the order type of the natural numbers.

[4] We know how big the continuum is—$2^{\aleph_0}$—but we cannot know whether that is as small as $\aleph_1$, or larger.

(g) "The roots of fifth-degree equations behave differently when interchanged, he [Galois] noticed, depending on whether or not they are all algebraic numbers. If they are, then interchanging them in the equation will result in another equation that is as algebraically sensible as the first one.... On the other hand, if some of the roots are transcendental numbers, interchanging them results in an algebraically nonsensical equation." (page 75)

(h) "Nim has an optimal strategy that is unfair because by following it, the first player will always win, and the second player will always lose."[5] (page 141)

(i) "And yet even with such assistance [computers], although it has helped us to formulate some successful chess strategies, we still have not been able to reason out the unique [?!] optimal strategy." (page 143)

(j) "For a nonpolynomial—or NP-class—problem, the necessary computer time grows exponentially with its size."[6] (page 177)

If that were not enough, the author also manages to give what would have been a correct, though not particularly illuminating, proof of the uncountability of the real numbers by the Cantor diagonal argument had he not said "irrational" whenever he meant to say "real." He later refers to the "middle" and "outer" points among three points on a circle—not so terrible, but he is using this as an example of a topological invariant!

Bridges to Infinity might make pleasant reading for the general public, but for math majors, it should be assigned with extreme caution.

REFERENCES

1. James C. Reber, Another required reading program for mathematics majors, this MONTHLY, 95 (1988) 867–868.
2. Michael Guillen, Bridges to Infinity, Jeremy P. Tarcher, Los Angeles, 1983.

Cauchy-Bunyakowski-Schwarz Inequality Revisited

FRANÇOIS DUBEAU
Département de mathématiques, Collège militaire royal de Saint-Jean, Saint-Jean-sur-Richelieu, Québec, Canada, J0J 1R0

Usually when we prove the Cauchy-Bunyakowski-Schwarz (CBS) inequality in any inner product space we consider the sign of a special quadratic polynomial, it is "a clever trick that is not easy to motivate" [**1**, p. 201]. It could be interesting to have other proofs of this result. For example in $\mathbb{R}^n$ (or $\mathbb{C}^n$) we can present a proof using the principle of mathematical induction.

Let $\vec{x} = (x_1, x_2, \ldots, x_n)$ and $\vec{y} = (y_1, y_2, \ldots, y_n)$ be any two vectors of $\mathbb{R}^n$. Let us consider the Euclidean inner product $\vec{x} \cdot \vec{y}$ defined by $\vec{x} \cdot \vec{y} = \sum_{i=1}^{n} x_i y_i$, and the Euclidean length of a vector defined by $\|\vec{x}\| = (\vec{x} \cdot \vec{x})^{1/2}$.

[5] Which player has the winning strategy depends on the initial position, and it is the position, and not the strategy, which is unfair.

[6] Then, by definition, $P \neq NP$!

CAUCHY-BUNYAKOWSKI-SCHWARZ INEQUALITY. *Let* $n \geqslant 1$ *and* $\vec{x}, \vec{y} \in \mathbb{R}^n$. *Then*

$$|\vec{x} \cdot \vec{y}| \leqslant \|\vec{x}\| \, \|\vec{y}\|. \tag{1}$$

and equality holds if and only if one vector is a scalar multiple of the other.

A proof of this result can be done using the upwards-downwards form of the principle of mathematical induction. This form of induction was used by A. Cauchy [**2**, pp. 375–377] to prove the arithmetic-geometric mean inequality. It can also be used to prove trigonometric inequalities [**3**, pp. 93–95].

Proof. Let $P(n)$ denote the proposition and let us start with $n = 2$.

Step 1. Proof that $P(2)$ *holds.* The result follows from

$$(x_1y_1 + x_2y_2)^2 + (x_1y_2 - x_2y_1)^2 = (x_1^2 + x_2^2)(y_1^2 + y_2^2).$$

Step 2. Proof that $P(n)$ *implies* $P(2n)$. With a pairing of indices two by two in the sum, we have

$$\begin{aligned}\left(\sum_{i=1}^{2n} x_i y_i\right)^2 &= \left[\sum_{j=1}^{n} (x_{2j-1}y_{2j-1} + x_{2j}y_{2j})\right]^2 \\ &\leqslant \left[\sum_{j=1}^{n} \left(x_{2j-1}^2 + x_{2j}^2\right)^{1/2}\left(y_{2j-1}^2 + y_{2j}^2\right)^{1/2}\right]^2 \\ &\leqslant \left(\sum_{i=1}^{2n} x_i^2\right)\left(\sum_{i=1}^{2n} y_i^2\right),\end{aligned}$$

where the first and second inequalities follow from $P(2)$ and $P(n)$, respectively.

Step 3. Proof that $P(n)$ *implies* $P(n-1)$. We only have to consider $x_n = y_n = 0$ for $P(n)$.

From these three steps we can conclude that (1) holds for $n \geqslant 1$. Equality follows from the condition for equality at step 1 and the arbitrary pairing in (2).

The first step of this proof is nothing but the case $n = 2$ of Lagrange's identity

$$\sum_{1 \leqslant i < j \leqslant n} (x_i y_j - x_j y_i)^2 + \left(\sum_{i=1}^{n} x_i y_i\right)^2 = \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right),$$

which can be proved by the upwards form of induction and can be used to prove the CBS-inequality [**2**, pp. 373–374].

Another proof of the CBS-inequality using the upwards form of induction can be done as follows: Step 1 as before, and

Step 2. Proof that $P(n)$ *implies* $P(n+1)$. We have

$$\begin{aligned}\left|\sum_{i=1}^{n+1} x_i y_i\right| &\leqslant \left|\sum_{i=1}^{n} x_i y_i\right| + |x_{n+1}y_{n+1}| \leqslant \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}\left(\sum_{i=1}^{n} y_i^2\right)^{1/2} \\ &\quad + \left(x_{n+1}^2\right)^{1/2}\left(y_{n+1}^2\right)^{1/2} \leqslant \left(\sum_{i=1}^{n+1} x_i^2\right)^{1/2}\left(\sum_{i=1}^{n+1} y_i^2\right)^{1/2},\end{aligned}$$

where the second and third inequalities follow from $P(n)$ and $P(2)$, respectively.

Once (1) is established for $\mathbb{R}^n$, we can extend it to more general inner product spaces. For example, let $w_1, w_2, \ldots, w_n$ be n strictly positive real numbers and consider the weighted Euclidean inner product $\vec{x} \cdot \vec{y} = \sum_{i=1}^{n} w_i x_i y_i$. Using the change of variables $\xi_i = \sqrt{w_i}\, x_i$ and $\eta_i = \sqrt{w_i}\, y_i$ we obtain (1) for any weighted Euclidean inner product space. More generally we can extend (1) to any finite dimensional inner product space using coordinates of vectors with respect to any orthonormal basis [**1**, p. 225].

Finally, for any inner product space we can establish (1) by a slight variation of the proof that projection onto an n-dimensional subspace is norm-decreasing [**4**, pp. 145–147]. Assume $\vec{x} \neq \vec{0}$ and let $\delta \in \mathbb{R}$ be such that $\vec{x} \cdot \vec{y} = \delta \|\vec{x}\| \, \|\vec{y}\|$. Let us remark that equality holds in (1) if and only if $|\delta| = 1$ or $\vec{y} = \vec{0}$. Let us consider the scalar λ such that $(\vec{y} - \lambda\vec{x}) \cdot \vec{x} = 0$. Hence $\lambda = \vec{x} \cdot \vec{y}/\|\vec{x}\|^2 = \delta\|\vec{y}\|/\|\vec{x}\|$ and $\|\vec{y} - \lambda\vec{x}\|^2 = (1 - \delta^2)\|\vec{y}\|^2$. It follows that $-1 \leqslant \delta \leqslant 1$ and $\vec{y} = \lambda\vec{x}$ if and only if $|\delta| = 1$ or $\vec{y} = \vec{0}$.

REFERENCES

1. H. Anton and C. Rorres, Elementary Linear Algebra with Applications, Wiley, New York, 1987.
2. A. Cauchy, Cours d'analyse de l'École Royale Polytechnique, Ire partie: analyse algébrique, Paris, Debure frères 1821, in Oeuvres complètes d'Augustin Cauchy, IIe série, tome III, Paris, Gauthier-Villars, 1897.
3. I. Niven, Maxima and Minima without Calculus, Dolciani Mathematical Expositions, No. 6, Mathematical Association of America, 1981.
4. G. Strang, Linear Algebra and Its Applications (3rd ed.), Harcourt, Brace, Jovanovich, San Diego, 1988.

An Old Max-Min Problem Revisited

MARY EMBRY-WARDROP
Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859

Recently, while teaching a beginning calculus class, I introduced the problem of inscribing a rectangle of maximum area in a given right triangle and asked the class how the rectangle should be oriented inside the triangle. I was slightly surprised when a student suggested that one side of the rectangle should be on the hypotenuse of the triangle. In previous classes the suggestions were to fit one corner of the rectangle into the right angle of the triangle. The class soon bogged down in trying to solve the problem with one side of the rectangle on the hypotenuse, but subsequently solved the problem using the other orientation.

I was sufficiently intrigued to solve the problem with both orientations and was surprised to discover that in either case the maximum area was $\frac{1}{4}\, bh$, where b and h are the base and height of the triangle. Furthermore in each case corners of the rectangle of maximum area lie on the midpoints of the two legs of the triangle.

At this point I was thoroughly intrigued: Could the rectangle be more advantageously oriented and for other orientations did a corner of the rectangle of maximum area lie on the midpoint of a leg of the triangle?

The following assertions describe the results of my investigation. The basic mathematical tools used were the Law of Sines and the technique of maximizing or

minimizing a function, usually discussed in a beginning calculus class. I omit the solutions under the assumption that most mathematicians enjoy finding their own solutions more than reading another's.

Let the rectangle be oriented in the triangle as pictured in FIGURE 1. The angle α satisfies $\theta \leqslant \alpha \leqslant \pi/2$. For a given angle α let $A(\alpha)$ be the area of the rectangle of maximum area which can be so inscribed in the triangle.

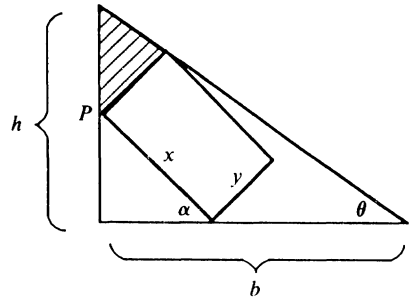


FIG. 1

Assertion 1. Given α the rectangle of largest area is obtained when

$$x = \frac{h}{2\sin\alpha} \quad \text{and} \quad y = \frac{b\sin\theta}{2\cos(\alpha - \theta)}.$$

In this case

$$A(\alpha) = \frac{bh\sin\theta}{4\sin\alpha\cos(\alpha - \theta)}$$

and P is the midpoint of the side of the triangle opposite the angle θ.

Assertion 2. The maximum area function $A(\alpha)$, $\theta \leqslant \alpha \leqslant \frac{\pi}{2}$, has largest value at the two endpoints of the interval:

$$A(\theta) = A\left(\frac{\pi}{2}\right) = \frac{1}{4}bh.$$

Assertion 3. The maximum area function $A(\alpha)$, $\theta \leqslant \alpha \leqslant \frac{\pi}{2}$ has smallest value at the midpoint of the interval:

$$A\left(\frac{1}{2}\theta + \frac{\pi}{4}\right) = \frac{bh\sin\theta}{2(1 + \sin\theta)}.$$

Moreover in this case, the shaded triangle in Figure 1 is isosceles with equal sides of length $\frac{1}{2}h$.

There are numerous problems here for a first-semester calculus class and each of the solutions calls upon skills that the students should have developed in geometry, trigonometry, and calculus. One of the interesting facets of Assertions 2 and 3 is that the students would be invited to consider maximizing and minimizing $A(\alpha)$, each value of which is itself a maximum. Few problems of this type seem to be accessible for beginning students. Furthermore, there is a collection of geometric

problems only touched upon in Assertions 1–3: Why is P a fixed point as a corner of the rectangles of maximum area? Why is the shaded triangle isosceles when $A(\alpha)$ is minimum? Why is $A(\alpha)$ minimum at the midpoint of the domain of $A(\alpha)$? Is there a geometric proof that $A(\alpha)$ is obtained with sides given by the equation in Assertion 1 or a geometric proof that $A(\theta)$ and $A(\frac{\pi}{2})$ are the maxima of $A(\alpha)$ and $A(\frac{1}{2}\theta + \frac{\pi}{4})$ is the minimum?

It is interesting to note that the problem of inscribing a rectangle of maximum area in an acute triangle is no more difficult than the same problem for a right triangle. Let the rectangle be oriented in an acute triangle, as pictured in FIGURE 2. The angle α satisfies $\theta_2 \leqslant \alpha \leqslant \pi/2$.

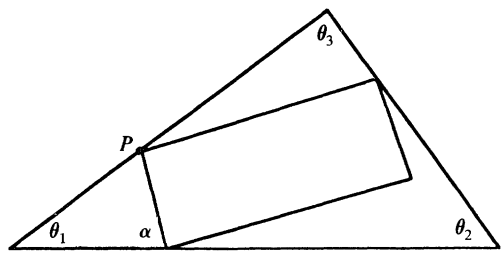


FIG. 2

Once again for a given angle α the rectangle of maximum area, $A(\alpha)$, which can be so inscribed is obtained when P is the midpoint of the side of the triangle opposite θ_2. The maximum of $A(\alpha)$ occurs at the endpoints θ_2 and $\frac{\pi}{2}$ and the minimum occurs at the midpoint $\frac{1}{2}\theta_2 + \frac{\pi}{4}$ of the domain of A. Finally, if the triangle is obtuse, $\theta_3 > \pi/2$ for example, the solution is the same except that $A(\alpha)$ would only be defined for $\frac{\pi}{2} - \theta_1 \leqslant \alpha \leqslant \frac{\pi}{2}$ and in this case the maximum of $A(\alpha)$ would be attained only at $\pi/2$.

Surely the results of this note are known. However, not a single mathematician to whom I have mentioned them admitted having considered the maximum area problem with any orientation other than the most standard one.

Functional Equations and L'Hôpital's Rule in an Exact Poisson Derivation

J. ACZÉL
Centre for Information Theory and Quantitative Economics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

In [2], Cooke derived the Poisson formula

$$P_n(t) = \text{Prob}[n \text{ arrivals in the interval } (0, t)]$$
$$= \frac{(\lambda t)^n}{n!} e^{-\lambda t} \qquad (\lambda \geqslant 0;\ n = 1, 2, \ldots) \tag{1}$$

from $P_0(t) = e^{-\lambda t}$, L'Hôpital's rule, the differentiability of P_n $(n = 1, 2, \dots)$, and the functional equations

$$P_n(t + h) = P_n(t)P_0(h) + P_{n-1}(t)P_1(h) \qquad (t, h > 0;\ n = 1, 2, \dots). \quad (2)$$

Actually, (1) does not satisfy (2), but in [2] equation (2) is regarded valid only "for very small h." There (2) is presented as a consequence of "independence and negligibility assumptions."

Here we state these assumptions explicitly, derive from them the exact functional equations in place of (2) and arrive at (1) in an exact and still elementary way, again by use of L'Hôpital's rule. The proof goes back at least 40 years; see [3] of which we give a somewhat modified form, more suitable for teaching beginners.

A (tacit) assumption is that of *homogeneity*: the probability $P_n(t_2 - t_1)$ of n occurrences in the interval (t_1, t_2) depends only on (n and) the length $t_2 - t_1$ of the interval. *Independence* means that the number of occurrences in (t_1, t_2) and (t_2, t_3) is independent. *Negligibility* (or *rarity*) means that in "very small" intervals *one* occurrence is approximately (asymptotically) equally probable to *at least one* occurrence; in terms of a formula:

$$\lim_{t \to 0} \frac{1 - P_0(t)}{P_1(t)} = 1, \quad (3)$$

that is,

$$\lim_{t \to 0} \frac{P_1(t) + P_2(t) + P_3(t) + \cdots}{P_1(t)} = 1.$$

Since the P_n are probabilities, we have

$$\sum_{n=0}^{\infty} P_n(t) = 1 \quad \text{and} \quad 0 \leqslant P_n(t) \leqslant 1 \qquad (n = 0, 1, 2, \dots). \quad (4)$$

Consequently,

$$\lim_{t \to 0} \frac{P_m(t)}{P_1(t)} = 0 \qquad (m = 2, 3, \dots). \quad (5)$$

By the independence assumption (and homogeneity),

$$P_n(t + u) = P_n(t)P_0(u) + P_{n-1}(t)P_1(t) + \cdots + P_0(t)P_n(u)$$
$$(t, u > 0;\ n = 0, 1, 2, \dots) \quad (6)$$

(there are exactly n occurrences in $(0, t + u)$ if there are either n occurrences in $(0, t)$ and none in $(t, t + u)$, or $(n - 1)$ in $(0, t)$ and one in $(t, t + u)$, ... or 0 in $(0, t)$ and n in $(t, t + u)$). This is the sequence of (exact) functional equations from which we will prove (1) with the aid of (3), (4) and (5).

We start with $n = 0$. Then (6) reduces to

$$P_0(t + u) = P_0(t)P_0(u) \qquad (t, u > 0). \quad (7)$$

The general positive, bounded solution of (7) is

$$P_0(t) = e^{-\lambda t} \quad (8)$$

with constant λ (see Appendix). By (4), $0 \leqslant P_0(t) \leqslant 1$, which shows that P_0 is bounded and that $\lambda \geqslant 0$. If positivity of P_0 is not required then $P_0(t) \equiv 0$ could be

a solution too, but considering (6) for $n = 1, 2, \ldots$ would then give $P_n(s) \equiv 0$ for $n = 1, 2, \ldots$, contrary to $\sum_{n=1}^{\infty} P_n(s) = 1$ in (4). Similarly, if $\lambda = 0$, $P_0(t) \equiv 1$, then from (4) $P_m(t) \equiv 0$ for $m = 1, 2, \ldots$ which is of the form (1) (if $(0 \cdot t)^0 = 1$), but uninteresting. So, in what follows, we take $\lambda > 0$.

Now we get to $n = 1$. In this case (6), combined with (8), gives

$$P_1(t + u) = P_1(t)e^{-\lambda u} + e^{-\lambda t}P_1(u) \qquad (t, u > 0).$$

If we multiply both sides by $e^{\lambda(t+u)}$ we see that f_1, defined by

$$f_1(t) = e^{\lambda t}P_1(t), \tag{9}$$

satisfies

$$f_1(t + u) = f_1(t) + f_1(u) \qquad (t, u > 0). \tag{10}$$

By (4) and (9), f_1 is also bounded on every finite interval. The bounded solution of (10) is

$$f_1(t) = c_1 t,$$

where c_1 is an arbitrary constant (see Appendix). By (9),

$$P_1(t) = c_1 te^{-\lambda t},$$

and by (3) and (8)

$$1 = \lim_{t \to 0} \frac{1 - e^{-\lambda t}}{c_1 te^{-\lambda t}} = \lim_{t \to 0} \frac{e^{\lambda t} - 1}{c_1 t} = \lim_{t \to 0} \frac{\lambda e^{\lambda t}}{c_1} = \frac{\lambda}{c_1}$$

(here we used the *L'Hôpital rule*—or the definition of the derivative of $e^{\lambda t}$ at 0), so that

$$P_1(t) = \lambda te^{-\lambda t}. \tag{11}$$

We finish the proof of (1) by induction. By (8) and (11) it is true for $n = 0$ and $n = 1$. Suppose (1) is true for $n = 0, 1, \ldots, m - 1$ (> 0). Then (6) for m reads

$$P_m(t + u) = P_m(t)e^{-\lambda u} + \lambda^m e^{-\lambda t}e^{-\lambda u}\left(\frac{t^{m-1}u}{(m-1)!1!} + \frac{t^{m-2}u^2}{(m-2)!2!} + \cdots + \frac{tu^{m-1}}{1!(m-1)!} \right) + e^{-\lambda t}P_m(u)$$

$$= P_m(t)e^{-\lambda u} + P_m(u)e^{-\lambda t} + \frac{\lambda^m e^{-\lambda t}e^{-\lambda u}}{m!}\left[(t + u)^m - t^m - u^m\right].$$

Multiplying this equation by $e^{\lambda(t+u)}$ and writing

$$f(t) = P_m(t)e^{\lambda t} - \frac{(\lambda t)^m}{m!}$$

gives again

$$f(t + u) = f(t) + f(u) \qquad (t > 0, u > 0). \tag{12}$$

With P_m also f is bounded on every finite interval and so (see Appendix), $f(t) = ct$

and

$$P_m(t) = \left(\frac{(\lambda t)^m}{m!} + ct \right) e^{-\lambda t}. \tag{13}$$

Applying also (5) and (11), we get

$$0 = \lim_{t \to 0} \frac{\left(\frac{(\lambda t)^m}{m!} + ct \right) e^{-\lambda t}}{\lambda t e^{-\lambda t}} = \lim_{t \to 0} \left(\frac{(\lambda t)^{m-1}}{m!} + \frac{c}{\lambda} \right) = \frac{c}{\lambda},$$

so that $c = 0$ and (13) shows that (1) holds also for m, which concludes the proof.

APPENDIX

Of course, if the P_n are differentiable, we immediately get from (7)

$$P_0'(t)P_0(u) = P_0'(t + u) = P_0(t)P_0'(u)$$

so that

$$P_0'(t) = \alpha P_0(t), \qquad P_0(t) = ae^{\alpha t}$$

and, substituting back into (7), $a = 0$ or $a = 1$, that is, $P_0(t) \equiv 0$ or $P_0(t) = e^{\alpha t}$. (From (4), as we have seen, $\alpha = -\lambda \leqslant 0$.)

The solution of (12) (and (10)) in this case is even simpler: $f'(t + u) = f'(t) =$ constant (say, c), so $f(t) = ct + b$ which satisfies (12) only if $b = 0$, $f(t) = ct$.

If, as above, we know only that f is bounded on at least one finite proper interval, say $(0, 1]$, (which followed from (4)) then $f(t) = ct$ is still the only solution of (12). We indicate here an elegant proof of Young [**4**]. Define

$$g(t) = f(t) - f(1)t. \tag{14}$$

This g too is bounded, say $A \leqslant g(t) \leqslant B$ on $(0, 1]$, satisfies the same equation

$$g(t + u) = g(t) + g(u) \tag{15}$$

and also $g(1) = 0$. So $g(t + 1) = g(t)$ for all $t > 0$ and

$$A \leqslant g(t) \leqslant B \quad \text{for all} \quad t > 0. \tag{16}$$

We prove that $g(t) \equiv 0$. Indeed, if there existed a $t_0 > 0$ with $g(t_0) = C \neq 0$, say $C > 0$, then by repeated application of (15) $g(nt_0) = ng(t_0) = nC > B$ for large enough n, contrary to (16). So $g(t) \equiv 0$ and, by (14), $f(t) = f(1)t = ct$ for all $t > 0$, as asserted.

As to (7), it is easy to see that P_0 is either everywhere or nowhere 0, thus either $P_0(t) \equiv 0$ or P_0 is positive for all $t > 0$, so that logarithms can be taken in order to reduce it to (12). (We know now only that $\log P_0(t)$ is bounded from above; $P_0 > 0$ gives no lower bound for $\log P_0$, but a slight modification of Young's argument (see [**1**]) gives the same result in this case too.)

REFERENCES

1. J. Aczél, On Applications and Theory of Functional Equations. Birkhäuser, Basel and Stuttgart, and Academic Press, New York, 1969.
2. W. P. Cooke, L'Hôpital's rule in a Poisson derivation, this MONTHLY, 95 (1988) 253–254.
3. L. Jánossy, A. Rényi and J. Aczél, On composed Poisson distributions, *I. Acta Math. Acad. Sci. Hung.*, 1 (1950) 209–224.
4. G. S. Young, The linear functional equation, this MONTHLY, 65 (1958) 37–38.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, FREDERICK W. LUTTMAN, MARVIN MARCUS, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, and EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

*An asterisk (*) after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before October 31, 1990 at the MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3384. *Proposed by Jacob Sturm, Rutgers University, Newark, NJ.*

Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of distinct complex numbers such that $\Sigma|a_n|$ converges. Let $\{c_n\}_{n=1}^{\infty}$ be a bounded sequence of non-zero complex numbers. Prove that there are infinitely many positive integers k such that

$$\sum_{n=1}^{\infty} c_n a_n^k \neq 0.$$

E 3385. *Proposed by Solomon W. Golomb, University of Southern California, Los Angeles.*

For what positive integers n is it true that

$$\sum_{p \leq \pi(n)} p = n?$$

Here, as usual, $\pi(x)$ denotes the number of primes not exceeding x and the summation is over primes p.

E 3386. *Proposed by Eugene F. Schuster, University of Texas, El Paso.*

Let L be the length of a $(2N - 1)$-game World Series, modeled as a sequence of independent identically distributed Bernoulli trials which terminates as soon as one team wins N games. (The length is the number of games actually played.) Prove the

seemingly obvious observation that the expected length $E(L)$ of the series is maximized when the two teams are evenly matched.

E 3387. *Proposed by Douglas B. West, University of Illinois at Urbana-Champaign.*

Determine the maximum possible number of edges in a simple graph on ten vertices that contains no 4-cycle and find all graphs achieving this maximum.

E 3388. *Proposed by Matthew Cook (student), University of Illinois at Urbana-Champaign, Walther Janous, Ursulinengymnasium, Innsbruck, Austria, and Marcin E. Kuczma, University of Warsaw, Poland.*

Let x_1 and x_2 be arbitrary positive numbers. Suppose we define a sequence $\{x_n\}_{n=1}^{\infty}$ by putting $x_{n+2} = 2/(x_{n+1} + x_n)$ for $n = 1, 2, 3, \ldots$. Prove that the sequence converges.

E 3389. *Proposed by Joseph H. Silverman, Brown University, Providence, RI.*

Let $\|t\|$ be the distance from the real number t to the nearest integer. Prove that there is a unique real number r greater than 1 such that

$$\sum_{n=1}^{\infty} r^{-2n}\|r^{n-1}t\|$$

is a polynomial in $\|t\|$; give this polynomial explicitly for the special value of r referred to.

E 3390. *Proposed by Robert B. Israel, University of British Columbia, Vancouver, Canada.*

Let $\vec{r}$, $\vec{v}$, and $\vec{a}$ be the position, velocity, and acceleration vectors of a particle at time t. Suppose the particle moves so that $\vec{a}$ is always perpendicular to both $\vec{r}$ and $\vec{v}$.

(a) Show that $\vec{v}_\infty = \lim_{t\to\infty} \vec{v}$ exists and show that $t\vec{v}_\infty - \vec{r}$ is bounded.

(b) Show that if $\int_0^\infty t|\vec{a}(t)|\,dt < \infty$, then $\lim_{t\to\infty}(t\vec{v}_\infty - \vec{r})$ exists.

SOLUTIONS OF ELEMENTARY PROBLEMS

A Combinatorial Sum

E 3290 [1988, 872]. *Proposed by Douglas B. West, University of Illinois, Urbana IL, and Douglas H. Wiedemann, Institute for Defense Analyses, Princeton, NJ.*

Given nonnegative integers l, m, n with $l, n \leqslant m$, evaluate the double sum

$$\sum_{i=0}^{l}\sum_{j=0}^{i}(-1)^j\binom{m-i}{m-l}\binom{n}{j}\binom{m-n}{i-j}.$$

Solution I by Irl C. Bivens, Davidson College, Davidson, NC. The value of the sum is $2^l\binom{m-n}{l}$. To see this, suppose that a mathematics department contains m

students, of whom n are male, and that all m students wish to take a class in combinatorics that is limited to l students and is graded pass/fail. There are $2^l\binom{m-n}{l}$ potential grade reports under the restriction that no males take the class. We show that the double sum also counts this.

The nonzero terms of the double sum have $0 \leqslant j \leqslant i \leqslant l$. Given such parameters, the number of grade reports in which i students pass, of whom j are male, is $\binom{n}{j}\binom{m-n}{i-j}\binom{m-i}{m-l}$: choose the males who pass, choose the females who pass, and from the remaining students choose those who do not take the course. Hence the double sum is the difference between the number of grade reports in which an even number of males pass and the number in which an odd number of males pass.

Furthermore, the operation of changing the grade of the eldest male taking the course establishes a bijection between the set S of grade reports in which an odd number of males pass and the set T of grade reports in which an even number of males pass *and* at least one male is in the class. This involution implies that the double sum is the number of grade reports for classes containing no males.

Solution II by Richard Stong, Harvard University, Cambridge, MA. The value of the sum is $2^l\binom{m-n}{l}$. Denoting the double sum by $S(l, m, n)$, we claim that $\sum_{l=0}^{m} S(l, m, n)x^l = (1 + 2x)^{m-n}$, from which the result follows by the binomial theorem.

Extending the sums over all nonnegative i and j does not change the value; we do so to facilitate interchange of summation. We compute the triple sum by viewing the sums in reverse order and applying the binomial theorem three times.

$$\begin{aligned}
\sum_l S(l, m, n)x^l &= \sum_j (-1)^j \binom{n}{j} \sum_i \binom{m-n}{i-j} x^i \sum_l \binom{m-i}{l-i} x^{l-i} \\
&= \sum_j (-1)^j \binom{n}{j} \sum_i \binom{m-n}{i-j} x^i (1+x)^{m-i} \\
&= (1+x)^m \sum_j (-1)^j \binom{n}{j} \left(\frac{x}{1+x}\right)^j \sum_i \binom{m-n}{i-j} \left(\frac{x}{1+x}\right)^{i-j} \\
&= (1+x)^m \sum_j (-1)^j \binom{n}{j} \left(\frac{x}{1+x}\right)^j \left(\frac{1+2x}{1+x}\right)^{m-n} \\
&= (1+2x)^{m-n} (1+x)^n \left(1 - \frac{x}{1+x}\right)^n \\
&= (1+2x)^{m-n}.
\end{aligned}$$

Editorial comment. Alternative solution techniques included everything from recurrence relations to calculus and complex analysis. David Callan and Fred Kochman observed that interchanging the order of summation gives an inner sum that is a special case of the binomial coefficient identity given as equation (24) in Donald E. Knuth, *The Art of Computer Programming* (vol. 1, second edition), p. 58, after which one application of the binomial theorem completes the sum.

Solved also by J. Anglesio (France), J. C. Binz (Switzerland), D. Borwein, D. Callan, J. Duemmel, A. J. Humphreys (student), A. A. Jagers (The Netherlands), N. L. Johnson, F. Kochman, C. Krattenthaler (Austria), Y. H. H. Kwong, H. Lipman, O. P. Lossers (The Netherlands), R. Martin (student, W. Germany), L. E. Mattics, A. Nijenhuis, T. S. Norfolk, R. Perez-Marco (Spain), H. Steinacker (Austria), J. H. Steelman, J. Tóth (Hungary), M. Vowe (Switzerland), A. Wender, and the proposers.

When the Kronecker Product is Triangular

E 3298 [1988, 954]. *Proposed by Marvin Marcus and Claire Pesce, University of California, Santa Barbara.*

If A and B are square matrices, we define the Kronecker product $A \otimes B$ as the matrix

$$\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nn}B \end{pmatrix},$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Let $A_1, A_2, \ldots, A_m$ be square matrices over a field, and let $K = A_1 \otimes A_2 \otimes \cdots \otimes A_m$. Prove that K is lower triangular if and only if the following two conditions are satisfied: (a) A_1 is lower triangular; (b) if there exists $p < m$ such that A_{p+1} is not lower triangular but all of $A_1, A_2, \ldots, A_p$ are lower triangular, then at least one of the matrices $A_1, A_2, \ldots, A_p$ has zero main diagonal.

Solution by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. It should be assumed that none of the matrices $A_1, A_2, \ldots, A_m$ is a zero matrix. With this assumption, we prove the result by induction on m. For any square matrices A and B, the entries of $A \otimes B$ that occur above the main diagonal are $a_{ij}b_{kl}$ with $i < j$ or with $i = j$ and $k < l$. If $B \neq 0$, then these will all be zero if and only if A is lower triangular and either B is also lower triangular or A has zero main diagonal. This verifies the statement for $m = 2$.

Suppose $m > 2$ and $K = A_1 \otimes A_2 \otimes \cdots \otimes A_m$ is lower triangular. Since the Kronecker product is associative, $K = A_1 \otimes (A_2 \otimes \cdots \otimes A_m)$. Hence the result for $m = 2$ implies that A_1 is lower triangular. Now suppose there exists $p < m$ such that A_{p+1} is not lower triangular but all of $A_1, A_2, \ldots, A_p$ are lower triangular. Then $K = A \otimes B$, where $A = A_1 \otimes \cdots \otimes A_p$ and $B = A_{p+1} \otimes \cdots \otimes A_m$. By the inductive hypothesis, B is not lower triangular. Therefore, A must have zero main diagonal. However, the main diagonal of A consists of all products formed using one main diagonal entry from each of $A_1, \ldots, A_p$. Hence one of these matrices must have zero main diagonal.

Conversely, suppose $m > 2$ and $K = A_1 \otimes (A_2 \otimes \cdots \otimes A_m)$ satisfies conditions (a) and (b). If A_2 is not lower triangular, then A_1 has zero main diagonal, and the result with $m = 2$ implies that K is lower triangular. If A_2 is lower triangular, then $A_2 \otimes \cdots \otimes A_m$ also satisfies (a) and (b) and by the inductive hypothesis is lower triangular. Now the case $m = 2$ implies that K is also lower triangular.

Solved also by J. K. Baksalary (Poland), D. Callan, Y. Ikeda, R. H. Jeurissen (The Netherlands), O. P. Lossers (The Netherlands), A. Nijenhuis, and the proposers. One incorrect solution was received.

Computing the Volume of a Zonotope

E 3301 [1989, 54]. *Proposed by Hugh L. Montgomery, University of Michigan, Ann Arbor, and Jeffrey D. Vaaler, University of Texas, Austin.* Suppose $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N$ are vectors in $\mathbb{R}^M$, where $N \geqslant M$. Prove that the set of vectors in $\mathbb{R}^M$ expressible in the form

$$\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \cdots + \lambda_N\mathbf{a}_N, \qquad 0 \leqslant \lambda_i \leqslant 1,$$

has M-dimensional volume

$$\sum_I |\det A_I|,$$

where I runs over the subsets of $\{1, 2, \ldots, N\}$ of cardinality M and A_I denotes the M by M matrix whose columns are the $\mathbf{a}_i$ for which $i \in I$.

Solution by H Guggenheimer, West Hempstead, NY. Given N compact convex sets K_i in $\mathbb{R}^M$, their *Minkowski sum* (or *vector sum*) is defined by

$$K_1 + \cdots + K_N = \{x^1 + \cdots + x^N : x^i \in K_i, i = 1, \ldots, N\}.$$

A formula of Minkowski computes the volume of the vector sum as

$$V(K_1 + \cdots + K_N) = \sum V(K_{i_1}, \ldots, K_{i_M}),$$

where the summands are Minkowski's mixed volumes, and the sum ranges over $1 \leqslant i_1, \ldots, i_M \leqslant N$. For these basic definitions and the results used below, see T. Bonnesen and W. Fenchel, *Theory of Convex Bodies*, translated from the German, BCS Associates, Moscow, Idaho (1987) or H. G. Eggleston, *Convexity*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 47, Cambridge University Press, 1958.

Let $A_i = \{\lambda\mathbf{a}_i : 0 \leqslant \lambda \leqslant 1\}$ be the line segment joining the origin to the point $\mathbf{a}_i$, for $i = 1, \ldots, N$. Then $A_1 + \cdots + A_N$ is the set (called a *zonotope*) described in the problem. $V(A_{i_1}, \ldots, A_{i_M}) = 0$ when two of the line segments that appear in the mixed volume are the same, and the mixed volume is invariant under the permutations of its arguments. Hence, in this case, the preceding sum takes the form

$$V(A_1 + \cdots + A_N) = M!\sum V(A_{i_1}, \ldots, A_{i_M}),$$

where the indices $\{i_1, \ldots, i_M\}$ range over all M-sets taken from $\{1, 2, \ldots, N\}$. This gives the required formula, because $M!V(A_{i_1}, \ldots, A_{i_M})$ is the volume of the parallelopiped spanned by the segments $A_{i_1}, \ldots, A_{i_M}$, and this volume is equal to $|\det A_I|$, where A_I is the matrix with columns $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_M}$.

Editorial comment. Richard Stanley pointed out that the formula appears in the paper of G. C. Shephard, "Combinatorial properties of associated zonotopes," *Can. J. Math.* 26 (1974), 302–321, where it is attributed to Peter McMullen. Most solvers proceeded by showing that the given set could be dissected into parallelopipeds, which was the result proved in Shephard's paper. The formula has been applied in P. McMullen, "Volumes of projections of unit cubes," *Bull. London Math. Soc.* 16 (1984) 278–280, and in G. D. Chakerian and P. Filliman, "The measures of the projections of a cube," *Studia Sci. Math. Hungar.* 21 (1986) 103–110.

Solved also by M. Golomb, L. R. King, O. P. Lossers (The Netherlands), A. Nijenhuis, and the proposers.

Averaging to Integers

E 3302 [1989, 54]. *Proposed by Jim Delany, California Polytechnic State University, San Luis Obispo.*

The mean and standard deviation of any 7 consecutive integers are both integers. What natural numbers greater than 1 share this property with 7?

Composite solution by Roger Douglass and Mike Breen, Alfred University, Alfred, NY, and Roger B. Eggleton, University of Brunei Darussalam, Borneo. We prove the following generalization:

The mean and standard deviation of an arithmetic sequence of n integers with common integral difference d are both integers if and only if $(n-1)d$ is even and n is of the form $[(2+\sqrt{3})^k/2]$ for k a non-negative integer.

The mean and standard deviation of $\{a, a+d, \ldots, a+(n-1)d\}$ are $a + d(n-1)/2$ and $[(n^2-1)/3]^{1/2}\, d/2$, respectively. When n is odd, the mean is integral, and the standard deviation is integral if and only if $[(n^2-1)/3]^{1/2}$ is integral, because it must be rational and therefore an integer (necessarily even since n^2-1 is divisible by 8). When n is even, the mean will be integral if and only if d is even, and then the standard deviation is integral if and only if $[(n^2-1)/3]^{1/2}$ is integral. In either case, $(n-1)d$ must be even. To complete the proof, we must describe the positive integers n for which there is an integer t with $n^2 - 1 = 3t^2$.

The equation $n^2 - 3t^2 = 1$ is a Pell equation whose solutions are well-known. Since $n = 2$ and $t = 1$ is the smallest positive solution, the theory of these equations implies that every positive integral solution (n_k, t_k) satisfies $n_k + t_k\sqrt{3} = (2+\sqrt{3})^k$, for some positive integer k. Furthermore, $n_k - t_k\sqrt{3} = (2-\sqrt{3})^k$, and therefore $n_k = [(2+\sqrt{3})^k + (2-\sqrt{3})^k]/2$. Since n_k is an integer and $0 < 2-\sqrt{3} < 1$, we have $n_k = [(2+\sqrt{3})^k/2]$, as claimed.

Editorial comment. When d is odd (for example, when $d = 1$), the requirement that $(n-1)d$ be even forces n to be odd and thus k to be even, say, $k = 2j$. Thus n has the form $[(7+4\sqrt{3})^j/2]$ when d is odd, where j is a non-negative integer.

Ninety-eight correct solutions, 2 "empirical" solutions and 3 incorrect solutions were received.

ADVANCED PROBLEMS

6631. *Proposed by Ronald Evans, University of California at San Diego and Peter Montgomery, Unisys and University of California at Los Angeles.*

We say that $f(z) = a_0 + a_1 z + \cdots + a_k z^k$ is a unimodal reciprocal polynomial with positive coefficients if

$$0 < a_0 = a_k < a_1 = a_{k-1} < a_2 = a_{k-2} < \cdots < a_{\lfloor k/2 \rfloor} = a_{\lfloor (k+1)/2 \rfloor}.$$

(i) If $z_1, z_2, \ldots, z_n$ are the nth roots of 1 and if $2\pi/n < t$, prove that

$$\prod_{|\arg z_j| > t} (z - z_j)$$

is a unimodal reciprocal polynomial with positive coefficients. Here $-\pi < \arg z \leqslant \pi$.

(ii) If $z_1, z_2, \ldots, z_n$ are the nth roots of -1 and if $\pi/n < t$, prove that

$$\prod_{|\arg z_j| > t} (z - z_j)$$

is a unimodal reciprocal polynomial with positive coefficients.

6632. *Proposed by Gilbert Muraz, Institut Fourier, Université de Grenoble I, St. Martin d'Héres, France, and Pawel Szeptycki and Fred Galvin, University of Kansas, Lawrence.*

Let E be a measurable subset of $\mathbb{R}$ modulo 1 having positive measure. For real t let N_t be the set of positive integers n such that nt modulo 1 is in E. Suppose $\{a_n\}_{n=1}^{\infty}$ is a sequence of positive real numbers such that $\sum a_n = \infty$. Prove that

$$\sum_{n \in N_t} a_n = \infty$$

for almost all t in $[0, 1]$.

6633. *Proposed by Horacio Porta, University of Illinois at Urbana-Champaign.*

Suppose we are given a large positive number N and the further information that $N = k!$ for some positive integer k. Show that we can determine k in at most $C \log\log\log N$ steps.

SOLUTION OF ADVANCED PROBLEMS

Fireworks in a Normed Linear Space

6546 [1987, 469]. *Proposed by Juan Arias-de-Reyna, Universidad de Sevilla, Spain.*

Suppose E is a separable infinite-dimensional, normed linear space. Construct a sequence $x_1, x_2, x_3, \ldots$ in E such that

(a) $\lim_{n \to \infty} \|x_n\| = +\infty$.

(b) Every x^{**} in E^{**} is a weak* cluster point of the sequence $\{x_n\}$.

Composite solution by the proposer, Mahlon M. Day, University of Illinois, Urbana, and N. Tenney Peck, University of Illinois, Urbana. Begin by observing that if a set D is weakly dense in a normed space E, it is weak* dense in E^{**} when E is regarded as a subspace of E^{**} (this is Goldstine's theorem).

Now for any positive integer n, let S_n be the set of points in E of norm n. Let the E_n be n-dimensional subspaces of E such that $E_n \subset E_{n+1}$ and $\bigcup_n E_n$ is norm-dense in E. Choose the F_n to be finite $1/n$-nets in $E_n \cap S_n$, and set $F = \bigcup_n F_n$. (Cf. [**1**].) The following result shows that the obvious enumeration of F will be a sequence $x_1, x_2, x_3, \ldots$ with the desired properties.

We shall show that F is weakly dense in E. Let x belong to E_k and let U be a weak neighborhood of x. We may assume that

$$U = \{ y : |f_i(y) - f_i(x)| < \varepsilon \text{ for } 1 \leqslant i \leqslant m \},$$

where each f_i belongs to E^*. Let

$$\mu = \max\|f_i\|$$

and

$$V = \left\{ y: |f_i(y) - f_i(x)| < \frac{\varepsilon}{2}, 1 \leqslant i \leqslant m \right\}.$$

Now choose n so that (1) $n > k$, (2) $n > \|x\|$, (3) $n > 2\mu/\varepsilon$, and (4) $E_n \cap V$ is unbounded. Then x belongs to V and $E_n \cap V$ is unbounded, so V meets $S_n \cap E_n$. If z belongs to the intersection $V \cap S_n \cap E_n$, there is a y in F_n such that

$$\|y - z\| < \frac{1}{n}.$$

Finally, for each i,

$$\begin{aligned} |f_i(x) - f_i(y)| &\leqslant |f_i(x) - f_i(z)| + |f_i(z) - f_i(y)| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

REFERENCE

1. M. M. Day, Normed Linear Spaces, Springer-Verlag, 1958, Ch. II, §5, (2), p. 41 or 1973, Ch. II, §5, (2)(a), p. 48.

No other solutions were received.

Some Diophantine Equations Not Solved by Fermat

6558 [1987, 884]. *Proposed by Barry Powell, Kirkland, Washington.*

Let $p > 3$ be prime. By a theorem of Faltings [*Invent. Math.* 73 (1983), 349–366] the Fermat equations

$$2x^p + y^p = z^p, \qquad 2^{p-1}x^p + y^p = z^p, \quad \text{and} \quad x^p + y^p = z^p$$

have only finitely many nontrivial integral solutions.

(a) Show there exists a positive integer N such that for q prime, $m \geqslant 1$, and $q^m > N$, the equation

$$q^m x^{2p} + y^{2p} = z^{2p}$$

has no positive integral solutions with x, y and z pairwise relatively prime.

(b) Show that for any positive integer $n \geqslant 2$, the Diophantine equation

$$x^6 - y^6 = z^n$$

has at most finitely many integral solutions with x and y relatively prime.

Solution by the proposer. The consequence of Faltings' theorem was misstated; the given Fermat equations have only finitely many *relatively prime* solutions.

Also, the solution would be more direct if the Fermat equations had read

$$2x^p + y^p = z^p, \qquad 2^{2p-1}x^p + y^p = z^p, \qquad \text{and} \quad x^p + y^p = z^p,$$

but of course we must deal with the equations as printed. For (a), let $\{x_i, y_i, z_i\}_{i=1}^{K}$

be an enumeration of all solutions to the given Fermat equations. We choose

$$N = 4 \max_{1 \leqslant i \leqslant K} \{|x_i|^p, |y_i|^p, |z_i|^p\}. \tag{1}$$

Suppose first that $q \neq 2$, and assume that there exist positive pairwise relatively prime integers x, y, and z such that $q^m x^{2p} = z^{2p} - y^{2p} = (z^p + y^p)(z^p - y^p)$. Since $\gcd(y, z) = 1$, we get that $\gcd(z^p + y^p, z^p - y^p) = 1$ or 2. Hence, there exist relatively prime positive integers a and b such that $x = ab$ or $2ab$ and at least one of the following holds:

$$q^m a^{2p} = z^p \pm y^p \quad \text{and} \quad b^{2p} = z^p \mp y^p, \tag{2}$$

$$2q^m a^{2p} = z^p \pm y^p \quad \text{and} \quad 2^{2p-1} b^{2p} = z^p \mp y^p, \tag{3}$$

or

$$2^{2p-1} q^m a^{2p} = z^p \pm y^p \quad \text{and} \quad 2b^{2p} = z^p \mp y^p. \tag{4}$$

Considering (1) and the equation involving a which holds above, we get that for $1 \leqslant i \leqslant K$,

$$2|z_i|^p \leqslant \tfrac{1}{2}N < \tfrac{1}{2}q^m \leqslant \tfrac{1}{2}q^m a^{2p} \leqslant |z^p \pm y^p| < 2|z|^p.$$

Thus, $|z_i| < |z|$ where z belongs to a solution (b, y, z) of one of the equations involving b above. But this contradicts the requirement that $(b^2, \pm y, z)$ (in the case of (2) or (4)) or $(2b^2, \pm y, z)$ (in the case of (3)) is some (x_i, y_i, z_i).

For the case $q = 2$ the argument is the same, but one should replace (3) and (4) in the list above with

$$2^{m-1} a^{2p} = z^p \pm y^p \quad \text{and} \quad 2b^{2p} = z^p \mp y^p.$$

This concludes the proof of (a).

For (b), there are no nontrivial solutions for $n = 3$ by Fermat's Last Theorem for exponent 3. The usual method of proof by infinite descent for exponent 3 can also be used to rule out nontrivial solutions for $n = 2$. So suppose there exist integers x, y, z with $(x, y) = 1$ for which $x^6 - y^6 = z^n$ with $n \geqslant 4$. Now exactly one of x, y, or z is even. If it is z, then without loss of generality

$$x^3 - y^3 = 2^{n-1} a^n \quad \text{and} \quad x^3 + y^3 = 2b^n \tag{5}$$

where $(a, b) = 1$ and b is odd. Make the substitutions $x + y = 2e$ and $x - y = 2f$. Then e and f are integers and $(e, f) = 1$. Also, a congruence argument modulo 4 shows that exactly one of e or f is even. We obtain

$$2f(3e^2 + f^2) = 2^{n-1} a^n \quad \text{and} \quad 2e(e^2 + 3f^2) = 2b^n. \tag{6}$$

Now either (i) $3 \nmid ef$ or (ii) $3 \mid ef$.

In case (i) it follows from (6) that there are integers α and β such that

$$f = 2^{n-2}\beta^n, \qquad e = \alpha^n. \tag{7}$$

Hence, by (6) again, there is an integer γ such that

$$3\alpha^{2n} + 2^{2n-4}\beta^{2n} = \gamma^n, \tag{8}$$

since $(3e^2 + f^2, f) = 1$. Now by Faltings' Theorem, equation (8) has only finitely many integer solutions $\{\alpha, \beta, \gamma\}$. Hence

$$\max(|\alpha|, |\beta|, |\gamma|) < M \tag{9}$$

for some positive integer M. Hence $|e|$ and $|f|$ are both less than $2^{n-2}M^n$ and we see that

$$\max(|x|, |y|) < 2^{n-1}M^n. \tag{10}$$

Hence, the original equation has only finitely many solutions x, y, z with $(x, y) = 1$, z even and $3 \nmid ef$.

For case (ii) let's assume $3|e$ (the case $3|f$ can be handled in a similar manner). By (6) it follows that there are integers α, β, γ such that

$$e = 3^{n-1}\alpha^n, \qquad f = 2^{n-2}\beta^n \tag{11}$$

and, by (6) again,

$$3^{2n-2}\alpha^{2n} + 3 \cdot 2^{2n-4}\beta^{2n} = 3\gamma^n \tag{12}$$

since $(e^2 + 3f^2, e) = 3$. Now by Faltings' theorem equation (12) has only finitely many solutions in integers, and this implies (as in (i)) that the original equation has only finitely many solutions with $(x, y) = 1$, z even, and $3|ef$.

It remains to consider the case of odd z. Since $(x, y) = 1$ there are relatively prime odd integers a and b such that

$$x^3 - y^3 = a^n \quad \text{and} \quad x^3 + y^3 = b^n. \tag{13}$$

The substitutions $e = x + y$ and $f = x - y$ yield

$$f(3e^2 + f^2) = 4a^n \quad \text{and} \quad e(e^2 + 3f^2) = 4b^n, \tag{14}$$

where clearly $(e, f) = 1$. We again distinguish the two cases (i) $3 \nmid ef$ and (ii) $3|ef$. These lead, respectively, to equations of the form

$$3\alpha^{2n} + \beta^{2n} = 4\gamma^n \quad \text{and} \quad 3^{2n-1}\alpha^{2n} + \beta^{2n} = 4\gamma^n, \tag{15}$$

and again the result follows from Faltings' Theorem.

Editoral comment. Of related interest is the paper by B. Powell and P. Ribenboim, "Note on a paper by M. Filaseta regarding Fermat's Last Theorem," *Ann. Univ. Turku*, Ser AI, 187 (1985).

The editors thank Michael Filaseta for considerable help in the editing and verification of this solution.

No other solutions were received.

Coverup Revealed in Hilbert Space

6577 [1988, 665]. *Proposed by B. Bagchi, G. Misra, and N. S. N. Sastry, Indian Statistical Institute, Calcutta.*

(a) Let H be an infinite-dimensional inner-product space. Suppose that finitely many closed balls cover the surface S of the unit ball B in H. Prove that these balls also cover the center of B.

(b) Does the above assertion remain valid if H is any infinite-dimensional, normed linear space?

Combined solution by (independently) the Florida State University Mail Room Problem Group, O. P. Lossers (The Netherlands), and John Henry Steelman,

Indiana University of Pennsylvania. The answer to (b), and hence to (a), is yes. If m balls $B_1, \ldots, B_m$ (or any m closed convex sets) cover the surface of the unit sphere of H, the same situation holds for any m-dimensional subspace of H. Inside this subspace, if B_1 fails to cover 0, there is a hyperplane P through 0 that does not intersect B_1. Thus the $m-1$ closed convex sets $B_2 \cap P, \ldots, B_m \cap P$ cover $S \cap P$, the unit sphere of an $(m-1)$-dimensional space. This step may be iterated; if $B_2, \ldots, B_{m-1}$ also fail to cover 0, then B_m will contain the surface of a one-dimensional unit sphere and hence 0 by convexity.

Editorial comment. Most solvers stayed in infinite-dimensional space and applied some variation of the Hahn-Banach theorem. A few (S. K. Chung (Hong Kong), Jesús Ferrer (Spain), Pei Yuan Wu (Taiwan)) based their solution on results about the weak topology. After his solution, Gerd Herzog (West Germany) remarked that Borsuk proved the following much deeper result (see, e.g., K. Deimling, *Nonlinear Functional Analysis*, Springer-Verlag, Berlin, 1985, p. 22). Let H be a normed linear space, $n \leqslant \dim H$, and $B_1, \ldots, B_n$ a cover of S such that $S \cap B_i$ is closed. Then for some x in S both x and $-x$ belong to the same B_j. In response to a query, Tenney Peck showed that in Hilbert space there is no positive r such that if finitely many closed balls cover the surface of the unit ball, then at least one of them contains the ball of radius r centered at the origin.

Solved also by Mahlon Day, K. P. Hart (The Netherlands), H. Hunziker and V. Mascioni (Switzerland), Miguel Lacruz, Reiner Martin (West Germany), Eero Posti (Finland), William H. Ruckle, and Thomas Starbird. Part (a) was solved by A. A. Jagers (The Netherlands) and the proposers.

An Application of the Three Square Theorem

6580 [1988, 774]. *Proposed by P. A. Batnik, University of Illinois at Urbana-Champaign.*

Prove that if n is an odd positive integer, there exist integers x_1, x_2, x_3, x_4 such that

$$n = x_1^2 + x_2^2 + x_3^2 + x_4^2, \qquad x_1 \geqslant \lfloor \sqrt{n} \rfloor - 1.$$

Editorial remark. As Godfrey Isaacs and Joseph Lehner pointed out to the editors, the published statement of the problem contained a serious typographical error, namely, the symbol $\geqslant$ was inadvertently replaced by the symbol $>$. As the case $n = 23$ shows (or, more generally, the case $n = 16k^2 + 7$ for $k = 1, 2, 3, \ldots$), the assertion of the problem is false with strict inequality.

Solution by J. L. Selfridge, Northern Illinois University, DeKalb. We prove the stronger result that if n is a positive integer not divisible by 16, then there exist integers x_1, x_2, x_3, x_4 satisfying

$$n = x_1^2 + x_2^2 + x_3^2 + x_4^2, \qquad x_1 \geqslant \lfloor \sqrt{n} \rfloor - 1. \qquad (*)$$

This amounts to showing that we may take $x_1 = \lfloor \sqrt{n} \rfloor - 1$ or $x_1 = \lfloor \sqrt{n} \rfloor$. We indicate our choice of x_1 from among these two alternatives by specifying the parity of x_1. We require the following classical result.

LEMMA. *A positive integer can be written as a sum of three squares of integers if and only if it cannot be expressed in the form* $4^a(8b + 7)$, *where* a *and* b *are non-negative integers.*

If $n \equiv 1 \pmod 4$, we choose x_1 to be the even number in the set $\{\lfloor\sqrt{n}\rfloor - 1, \lfloor\sqrt{n}\rfloor\}$. Then $n - x_1^2 \equiv 1 \pmod 4$ and so $n - x_1^2$ is a sum of three squares by the above lemma.

If $n \equiv 2, 3, 4, 6, 7 \pmod 8$, we choose x_1 to be the odd number in the set $\{\lfloor\sqrt{n}\rfloor - 1, \lfloor\sqrt{n}\rfloor\}$. Then $n - x_1^2 \equiv 1, 2, 3, 5, 6 \pmod 8$ and so again $n - x_1^2$ is a sum of three squares by the above lemma.

If $n \equiv 8 \pmod{16}$, we choose x_1 to be the even number in the set $\{\lfloor\sqrt{n}\rfloor - 1, \lfloor\sqrt{n}\rfloor\}$. Then $n - x_1^2 \equiv 4, 8 \pmod{16}$ and so once more $n - x_1^2$ is a sum of three squares by the above lemma.

In all three of these cases there exist integers x_1, x_2, x_3, x_4 such that $(*)$ holds.

The assertion of the problem is not in general true for integers n divisible by 16; for example, it fails when $n = 1136$, since $1136 - 33^2 = 47$ and $1136 - 32^2 = 4^2 \cdot 7$. Further, if $n = 2^{2r+1}$, where $r \geqslant 3$, the only way of expressing n as a sum of four squares is (up to permutations)

$$n = (2^r)^2 + (2^r)^2 + 0^2 + 0^2 = \left(\sqrt{n/2}\right)^2 + \left(\sqrt{n/2}\right)^2 + 0^2 + 0^2,$$

where $2^r = \sqrt{n/2} \leqslant \lfloor\sqrt{n}\rfloor - 3$.

Solved also by David Callan, R. J. Chapman (England), M. Chiebik (Czechoslovakia), Carl Cotner and San Ling (students), O. P. Lossers (The Netherlands), Nick MacKinnon (England), and Carl Pomerance.

An Integral for Jacobi

6582 [1988, 774]. *Proposed by N. J. Fine, Deerfield Beach, FL.*

Prove that

$$\int_0^{e^{-\pi}} \prod_{n=1}^{\infty} \frac{(1 - q^{2n})^{20}}{(1 - q^n)^{16}}\, dq = \frac{1}{16}.$$

Solution by A. A. Jagers, Universiteit Twente, Enschede, The Netherlands. We start with

$$\sum_{n=-\infty}^{\infty} q^{n^2} = \prod_{n=1}^{\infty} (1 - q^{2n})(1 + q^{2n-1})^2,$$

a special case of Jacobi's triple product identity [Theorem 352 of the fifth edition of *An Introduction to the Theory of Numbers* by G. H. Hardy and E. M. Wright]. Since

$$\prod_{n=1}^{\infty} (1 + q^{2n-1}) = \prod_{n=1}^{\infty} (1 + q^n)(1 + q^{2n})^{-1}$$

$$= \prod_{n=1}^{\infty} \frac{(1 - q^{2n})(1 - q^{2n})}{(1 - q^n)(1 - q^{4n})},$$

we have

$$\left(\sum_{n=-\infty}^{\infty} q^{n^2}\right)^4 = \prod_{n=1}^{\infty} \frac{(1-q^{2n})^{20}}{(1-q^n)^8(1-q^{4n})^8}.$$

On the other hand,

$$\left(\sum_{n=-\infty}^{\infty} q^{n^2}\right)^4 = 1 + 8 \sum_{\substack{m=1 \\ m \not\equiv 0 \bmod 4}}^{\infty} \frac{mq^m}{1-q^m}$$

by Theorem 385 of Hardy and Wright, and logarithmic differentiation yields

$$\frac{d}{dq}\left\{q\prod_{n=1}^{\infty}\frac{(1-q^{4n})^8}{(1-q^n)^8}\right\} = \prod_{n=1}^{\infty}\frac{(1-q^{4n})^8}{(1-q^n)^8} \times\left\{1 + 8\sum_{m=1}^{\infty}\frac{mq^m}{1-q^m} - 8\sum_{m=1}^{\infty}\frac{4mq^{4m}}{1-q^{4m}}\right\}.$$

Hence,

$$\frac{d}{dq}\left\{q\prod_{n=1}^{\infty}\frac{(1-q^{4n})^8}{(1-q^n)^8}\right\} = \prod_{n=1}^{\infty}\frac{(1-q^{2n})^{20}}{(1-q^n)^{16}}.$$

Now recall that the Dedekind eta-function $\eta(\tau)$ is defined by

$$\eta(\tau) = q^{1/24}\prod_{n=1}^{\infty}(1-q^n),$$

where $q = \exp(2\pi i\tau)$ and $\operatorname{Im}(\tau) > 0$. Thus

$$\int_0^{e^{-\pi}} \prod_{n=1}^{\infty}\frac{(1-q^{2n})^{20}}{(1-q^n)^{16}}\,dq = e^{-\pi}\left\{\prod_{n=1}^{\infty}\frac{1-e^{-4\pi n}}{1-e^{-\pi n}}\right\}^8 = \left\{\frac{\eta(2i)}{\eta(-1/(2i))}\right\}^8 = \frac{1}{16},$$

since $\eta(-1/\tau) = \sqrt{-i\tau}\,\eta(\tau)$.

Editorial comment. Jagers' solution is essentially the same as the proposer's. Jesse Deutsch remarks that the integral appears in Nathan Fine's book *Basic Hypergeometric Series and Applications*, Mathematical Surveys and Monographs, No. 27, American Mathematical Society, on p. 89 (equation 33.43). Both Kee-Wai Lau (Hong Kong) and Klaus Zacharias (Berlin) presented solutions based on elliptic integrals and the Legendre relation

$$E(k)K(k') + E(k')K(k) - K(k)K(k') = \frac{\pi}{2}.$$

The solutions mentioned were the only ones received.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

The Mathematics of Plato's Academy. A New Reconstruction. By D. H. Fowler, Oxford Science Publications, the Clarendon Press Oxford University Press, New York, 1987, xxi + 401 pp.

RICHARD K. GUY
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada, T2N 1N4

In each of FIGURES 1–3, a side AB of unit length of a regular polygon is folded onto the part AE of a diagonal of length x (x is different in each figure, of course) forming two similar triangles ACF and CDE. FIGURE 1 shows the relationship between what we now call golden section (a name apparently coined by Martin Ohm, younger brother of the physicist, around 1834) and the regular pentagon (Euclid's *Elements*: II, 10; IV, 10 and 11). In fact,

$$\frac{DE}{CE} = \frac{CF}{AF},$$

$$\frac{x-1}{1} = \frac{1}{x}\ (*)\ x = 1 + \frac{1}{x} = 1 + \frac{1}{1 + \frac{1}{x}} = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \cdots}}}}.$$

In each of FIGURES 2 and 3,

$$\frac{AC}{CF} = \frac{CD}{DE}, \qquad \frac{x}{1} = \frac{1-y}{y}.$$

Add

$$\frac{1}{1} = \frac{y}{y}$$

to the two sides (componendo):

$$\frac{1+x}{1} = \frac{1}{y}, \qquad \frac{y}{1} = \frac{1}{1+x}.$$

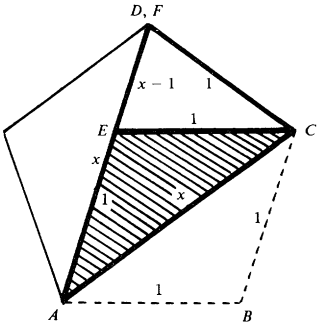


FIG. 1

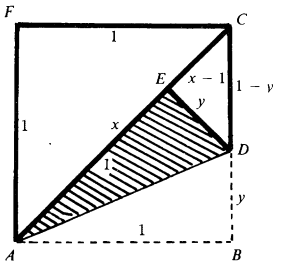


FIG. 2

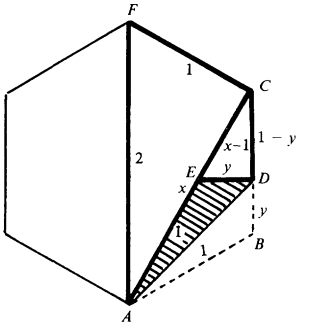


FIG. 3

In FIGURE 2,

$$x - 1 = y, \qquad x = 1 + y = 1 + \frac{1}{1+x} \qquad (*)$$

$$x = 1 + \frac{1}{2 + \dfrac{1}{1+x}} = 1 + \frac{1}{2 + \dfrac{1}{2 + \dfrac{1}{1+x}}} = 1 + \frac{1}{2 + \dfrac{1}{2 + \dfrac{1}{2 + \dfrac{1}{2 + \cdots}}}}.$$

In FIGURE 3,

$$\frac{x-1}{y} = \frac{2}{1}, \qquad x = 1 + 2y, \qquad \frac{1+x}{2} = 1 + y \qquad (*)$$

$$x = 1 + 2y = 1 + \frac{2}{1+x} = 1 + \frac{1}{1+y} = 1 + \frac{1}{1 + \dfrac{1}{1+x}} = 1 + \frac{1}{1 + \dfrac{1}{2 + \dfrac{1}{1+y}}}$$

$$= 1 + \frac{1}{1 + \dfrac{1}{2 + \dfrac{1}{1 + \dfrac{1}{1+x}}}} = 1 + \frac{1}{1 + \dfrac{1}{2 + \dfrac{1}{1 + \dfrac{1}{2 + \dfrac{1}{1+y}}}}}$$

$$= 1 + \frac{1}{1 + \dfrac{1}{2 + \dfrac{1}{1 + \dfrac{1}{2 + \dfrac{1}{1 + \dfrac{1}{2 + \cdots}}}}}}.$$

In present day language, we have found the continued fraction expansion for the golden mean, $(1 + \sqrt{5})/2$, for $\sqrt{2}$, and for $\sqrt{3}$. Each of the three arguments, at least as far as (*), have used only the idea of the ratio (*logos*) of lengths of line segments, very familiar to the Greeks. The claim is that the Greeks could also well understand the **anthyphairetic ratios** that we now call continued fractions. The author derives the name **anthyphairesis** from the Greek verb *anthuphairein*, which Euclid uses at *Elements* VII, 1 and 2, and X 2 and 3, for the operation of reciprocal subtraction, whereby he finds the greatest common measure: the process we now call the Euclidean algorithm, or more exactly, algorism, as the author prefers. In fact he suggests the ancient to modern "proportion":

(anthyphairesis) : (algorisms) : : (continued fractions) : (algorithms)

If we look at the process in one direction, it is not difficult to see how the Greeks discovered the incommensurability of $\sqrt{2}$ and other surds. In the other direction, it is not unduly fanciful to visualize the infinite process which generates ever closer rational approximations, while never attaining the limit. Compare several recent notes [**3**, **13**, **14**, **15**].

How did the Sumerians, a millenium before Pythagoras, obtain the remarkable approximation

$$1\frac{24}{60}\frac{51}{60^2}\frac{10}{60^3}$$

to $\sqrt{2}$? If the side of the square in Yale Babylonian Collection 7289 [**9**, Plate 6) were a mile, then, to mix metraphors, the error in its diagonal would be less than a tenth of a millimetre! Since when has man been able to measure as accurately as that?

Perhaps the Sumerians used the "Heronian algorithm," which we now recognize as a simple application of the Newton-Raphson process. One is too small, and two too big, so try their average, 3/2. On dividing into two, 4/3 is too small and 3/2 is too big, so try their average:

$$\frac{3^2 + 2 \cdot 2^2}{2 \cdot 2 \cdot 3} = \frac{17}{12}. \qquad \text{Try again:} \quad \frac{17^2 + 2 \cdot 12^2}{2 \cdot 12 \cdot 17} = \frac{577}{408}.$$

Not as good as the Sumerians, since this answer is nearer to 11 in the third sexagesimal place. But they must have noticed that the arithmetic mean was greater than the geometric mean, and that their approximations, after the first, were over-estimates, and they had the good judgement to take $10/60^3$ instead of $11/60^3$.

If we look at the convergents

$$\left(\frac{0}{1}\ \frac{1}{1}\right)\frac{1}{1}\ \frac{3}{2}\ \frac{7}{5}\ \frac{17}{12}\ \frac{41}{29}\ \frac{99}{70}\ \frac{239}{169}\ \frac{577}{408}\ \cdots$$

to the continued fraction we obtained near the outset, we see the approximations as the first, second, fourth and eighth [compare **2**, **4**].

This is not necessarily to suggest that the Sumerians had the concept of continued fractions, but the Greeks *could* have had.

The Babylonian astronomers found (but with not quite this accuracy!) that the seasons recur every 365.242199 days and that the period of the moon's phases is 29.530588 days. The ratio of these numbers gives the continued fraction

$$12 + \frac{1}{2 + \frac{1}{1 + \frac{1}{2 + \frac{1}{1 + \frac{1}{1 + \frac{1}{17 + \frac{1}{2}}}}}}}$$

as far as (or slightly further than) even modern astronomical measurements justify our going. The convergents

$$\frac{12}{1}\ \frac{25}{2}\ \frac{37}{3}\ \frac{99}{8}\ \frac{136}{11}\ \frac{235}{19}\ \frac{4131}{334}\ \frac{8497}{687}$$

are roughly the number of "months" in a "year." The ratio 235 : 19, found by truncating the large partial quotient 17, is a very good approximation, and is called the **Metonic cycle** since it was known to the Athenian astronomer Meton, who flourished a little earlier than Plato. It was almost certainly also known to earlier Babylonian astronomers and it is still used to determine both the Jewish calendar and the date of Easter.

The use of continued fractions to calculate gear ratios is now well known. But the corresponding results were known to, and implemented by, the Greeks [**11**, **12**].

Fowler invents illustrative conversations between a putative slave-boy of remarkable intelligence and prolonged youth, and each of Socrates, Eudoxus and Archytas, in which various aspects of anthyphairesis are developed, including the relationship with music theory. It is a pity that this has to be one of the most speculative areas as far as our knowledge is concerned. The possibility is explored of compounding $\sqrt{2}$: $1 = [1, 2, 2, 2, \ldots]$ and $\sqrt{3}$: $1 = [1, 1, 2, 1, 2, 1, 2, \ldots]$ to give $\sqrt{6}$: $1 = [2, 2, 4, 2, 4, 2, 4, \ldots]$, although even the slave-boy is unable to explain what is going on. Indeed, had he tried to *add* these ratios, the results would have been even more bewildering, and Fowler quotes Khinchin [**6**, p. 20] as Gosper had done before him:

> There is, however, another and yet more significant practical demand that the apparatus of continued fractions does not satisfy at all. Knowing the representations of several numbers, we would like to be able, with relative ease, to find the representations of the simpler functions of these numbers (especially, their sum and product). In brief, for an apparatus to be suitable from a practical standpoint, it must admit sufficiently simple rules for arithmetical operations; otherwise, it cannot serve as a tool for calculation. We know how convenient systematic fractions are in this respect. On the other

> hand, for continued fractions there are no practically applicable rules for arithmetical operations; even the problem of finding the continued fraction for a sum from the continued fraction representing the addends is exceedingly complicated, and unworkable in computational practice.

It is not as well known as it should be that Bill Gosper has in fact met this "significant practical demand." Some indications of this are found in [**1**] and in [**7**] and [**8**], but not in the book under review.

There are more misprints and inaccuracies than one expects from Clarendon Press. However, it is a scholarly and well-documented work. On the other hand it is highly idiosyncratic and many will find it over fanciful. At least the author warns us at the outset by quoting Neugebauer's closing paragraph [**9**, p. 177]:

> In the "Cloisters" of the Metropolitan Museum in New York there hangs a magnificent tapestry which tells the tale of the Unicorn. At the end we see the miraculous animal captured, gracefully resigned to his fate, standing in an enclosure surrounded by a neat little fence. This picture may serve as a simile for what we have attempted here. We have artfully erected from small bits of evidence the fence inside which we hope to have enclosed what may appear as a possible, living creature. Reality, however, may be vastly different from the product of our imagination; perhaps it is vain to hope for anything more than a picture which is pleasing to the constructive mind when we try to restore the past.

REFERENCES

1. R. William Gosper, Item 101 in Hakmem, AIM 239, M.I.T., Feb. 1972, 37–44.
2. M. A. Grant, Approximating square roots, Note 66.36, *Math. Gaz.* 66 (1982) 230–231.
3. Michael Charles Irwin, Geometry of continued fractions, this MONTHLY, 96 (1989) 696–703.
4. Kenneth R. Johnson, An iterative method for approximating square roots, *Math. Mag.* 62 (1989) 253–259.
5. W. Burton Jones and Wolfgang J. Thron, Continued Fractions, Encyclopedia of Mathematics and its Applications, 11, Addison-Wesley, 1980.
6. Alexander Ya. Khinchin, Continued Fractions, Univ. of Chicago Press, 1964.
7. Peter Kornerup and David W. Matula, An on-line arithmetic unit for bit-pipelined rational arithmetic, *J. Parallel Distrib. Comput.* 5 (1988) 310–330.
8. Peter Kornerup and David W. Matula, Finite precision lexicographic continued fraction number systems, *Proc. 7th IEEE Sympos. Comput. Arith.*, (1985) 207–214.
9. Otto Neugebauer, The Exact Sciences in Antiquity, Dover Publications, 1969.
10. Oskar Perron, Die Lehre von den Kettenbrüche, Teubner, Stuttgart, 1954.
11. Derek J. de Solla Price, The Equatorie of the Planetis, Cambridge Univ. Press, 1955, pp. 93–117.
12. Derek J. de Solla Price, Gears from the Greeks: the antikythera mechanism—a calendar computer from ca. 80 B.C., *Trans. Amer. Philos. Soc.* (N.S.) 64 Part 7 (1974).
13. Y. Sagher, What Pythagoras could have done, this MONTHLY, 95 (1988) 117.
14. David J. Sprows, Irrationals and the fundamental theorem of arithmetic, this MONTHLY, 96 (1989) 732.
15. William C. Waterhouse, Why square roots are irrational, this MONTHLY, 93 (1986) 213–214.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

Elementary, T(13). *Primer for College Algebra.* Karl J. Smith, Patrick J. Boyle. Brooks/Cole, 1990, xii + 563 pp. [ISBN: 0-534-12288-4] Excellent text "with the goal of preparing mature beginning students with the ...skills for college algebra." Also appropriate for students needing review. Good explanations and pointers in worked examples. Has diagnostic tests prior to each chapter. Not enough graphs. From beginning concepts to conic sections, exponents and logs, sequences, series, and the binomial theorem. FR

Precalculus, T(13: 1), S. *Precalculus: Functions and Graphs, Sixth Edition.* Earl W. Swokowski. PWS-Kent, 1990, xvi + 676 pp. [ISBN: 0-534-92086-1] Changes from the *Fifth Edition* (TR, June-July 1987): solutions to several examples have been rewritten; more applied problems have been added; and exercises have been expanded. SP

History, S*(15-16), P*, L*. *The History of Modern Mathematics.* Eds: David E. Rowe, John McCleary. Academic Pr, 1989. *Volume I: Ideas and Their Reception,* xvi + 453 pp, $39.50 [ISBN: 0-12-599661-6]; *Volume II: Institutions and Applications,* xvi + 325 pp, $37.50. [ISBN: 0-12-599662-4] Twenty-four essays on the mathematics of the 19th and 20th centuries. *Volume I* focuses on the emergence and reception of new ideas in pure mathematics, e.g., three articles on the assimilation and expansion of ideas of Gauss and Riemann. *Volume II* is concerned with ideas from mathematical physics and applied mathematics. Representative articles are the work of Liouville using Gauss' intrinsic approach to surfaces to derive the Hamilton-Jacobi equations, and the revolutionary impact of high-speed computers on numerical analysis through the work of von Neumann. Very interesting reading. MPR

Logic, T(14), L.** *Introduction to Proof in Abstract Mathematics.* Andrew Wohlgemuth. Saunders College, 1990, xii + 366 pp. [ISBN: 0-03-026782-X] Very interesting text attempting to ease the transition to abstract mathematics for the average student. Focuses on proof structures, techniques, and ideas. Concepts we assume the students just "pick up" along the way are clearly presented in simple contexts. Topic selection is excellent: sets and rules of inference; functions; relations; operations and integers (including an introduction to groups); proofs in analysis; cardinality; and groups. Truth tables are mercifully relegated to an appendix. Plenty of exercises and examples. Highly recommended for transition courses. MPR

Graph Theory, P, L. *Labyrinth Problems.* Armin Hemmerling. Teubner-Texte zur Math., Band 114. BG Teubner Leipzig, 1989, 215 pp, 27 DM (P). [ISBN: 3-322-00704-9] There are two classical problems in labyrinth theory. The first is the search problem: to find an algorithm guaranteed to eventually reach every point in the labyrinth. The second, clearly a special case of the first, is the escape problem. Interest in these problems is motivated by the fact that the possible configurations of a nondeterministic Turing machine can be modeled as a network. This book contains the main results, techniques, and open questions of labyrinth theory. SM

Graph Theory, P. *Algorithms On Graphs.* H.T. Lau. TAB Books, 1989, vi + 231 pp, $29.95. [ISBN: 0-8306-3429-0] Explicit Fortran programs for 20 problems in graph theory, including connectivity, shortest paths, traversability, node coloring, and planarity testing. From the Preface: "The main objective of this book is to provide computer programs that can be used with minimal effort for problem solving without much concern for their underlying methodology and implementation." BC

Graph Theory, P. *Coloring Theories.* Steve Fisk. Contemp. Math., V. 103. AMS, 1989, xv + 164 pp, $31 (P). [ISBN: 0-8218-5109-8] Global properties of colorings and maps of simplicial complexes. BC

Combinatorics, T(14-17: 1, 2), L**.** *Concrete Mathematics: A Foundation for Computer Science.* Ronald L. Graham, Donald E. Knuth, Oren Patashnik. Addison-Wesley, 1989, xiii + 625 pp. [ISBN: 0-201-14236-8] "What do Graham, Knuth, and Patashnik know that I don't?" Here is a good partial answer in an incredibly well-written presen-

tation of the sort of material that lives in Knuth's mathematical foundations in *The Art of Computer Programming*. This book represents nearly twenty years of classroom experience with material rooted in many fields: discrete mathematics, an antidote to the excesses of the "new math," and the need for bringing together the spirit of mathematics which inspires those neat algorithms some computer scientists use. Presented with grace and (marginal) graffiti. Contains among other things chapters on recurrent problems, sums, integer functions, special numbers (Stirling and the like), and asymptotics. Each chapter has many exercises (with answers) running from warmups through occasional research problems. JAS

Combinatorics, T*(13-17: 1, 2), S, L. *Introductory Combinatorics, Second Edition.* Kenneth P. Bogart. Harcourt Brace Jovanovich, 1990, xvii + 622 pp, $38. [ISBN: 0-15-541576-X] In addition to a number of improvements in clarity and style, there is new material throughout (e.g., more algorithmic content related to trees, matchings, block designs, coding, interval orders), and a new chapter on counting under the action of a group. Additional exercises and examples, and answers to odd numbered exercises. (*First Edition,* TR, May 1984.) LCL

Discrete Mathematics, T(14-15: 1), L. *Discrete Mathematics, Second Edition.* Richard Johnsonbaugh. Macmillan, 1990, xiv + 705 pp. [ISBN: 0-02-359690-2] Features extensive revision and expansion (from 349 pages to 582 pages) over the original edition (Extended Review, June-July 1987). Begins with new chapters on logic and proof, the language of mathematics, and algorithms; many changes and reorganization in later chapters as well as added examples and exercises. Appendix, references, hints and solutions to exercises. JS

Number Theory, P. *Automorphic Forms, Shimura Varieties, and L-functions.* Eds: Laurent Clozel, James S. Milne. Academic Pr, 1990, $49.50 each. *Volume I,* Perspectives in Math., V. 10, xii + 437 pp [ISBN: 0-12-176651-9]; *Volume II,* Perspectives in Math., V. 11, xii + 391 pp. [ISBN: 0-12-176652-7] Expanded lectures from a July 1988 conference held at the University of Michigan. May be viewed as an update of the proceedings of the 1977 AMS Summer Research Institute at Corvallis. Articles tend to be advanced surveys rather than focused on one individual's work. GG

Number Theory, P*. *Transcendental Numbers.* Andrei Borisovich Shidlovskii. Stud. in Math., V. 12. Transl: Neal Koblitz. Walter de Gruyter, 1989, xix + 466 pp, $89.50. [ISBN: 3-11-011568-9] Definitive exposition of the author's extensive work on transcendental numbers. A major focus is the transcendental nature of values of E-functions, a particular class of entire function. Includes a long bibliography and some supplementary remarks written for the English edition. A valuable reference book. SG

Linear Algebra, T(13-14: 1, 2). *Elementary Linear Algebra with Applications, Second Edition.* W. Keith Nicholson. PWS-Kent, 1990, xvi + 576 pp, $34. [ISBN: 0-534-92189-2] Changes from the *First Edition* (TR, February 1987) include highlighted definitions, new material on Euclidean n-space, some new exercises and examples, new sections on LU-factorization and complex matrices, and some rewriting and reorganization. RH

Algebra, T(15-17: 1, 2), L. *Abstract Algebra, An Introduction.* Thomas W. Hungerford. Saunders College, 1990, xxi + 575 pp, $49.25. [ISBN: 0-03-030558-6] Departs markedly from the pattern of most recent texts. Part 1: The Core Course, begins with the ring of integers through modular arithmetic, introduces abstract rings, and moves on to polynomial rings and congruences, ideals, and quotient rings. Groups are introduced in Chapter 7. Part 2: Advanced Topics, treats field extensions, more group theory, and Galois theory including solvability by radicals. Part 3: Excursions and Applications, has a selection of applications such as geometric constructions and coding theory. Appendices, bibliography, answers and hints for exercises. JS

Differential Equations, T(16-18), S, L. *Elementary Stability and Bifurcation Theory, Second Edition.* Gérard Iooss, Daniel D. Joseph. Undergrad. Texts in Math. Springer-Verlag, 1990, xxiii + 324 pp, $49.95. [ISBN: 0-387-97068-1] A substantially revised edition with efforts towards simplification and clarification, especially of material found in the final two chapters of the earlier version. An additional chapter on stability and bifurcation in conservative systems is included. KS

Differential Equations, P. *Lecture Notes in Control and Information Sciences-126: Stochastic Differential Systems.* Eds: N. Christopeit, K. Helmes, M. Kohlmann. Springer-Verlag, 1989, ix + 342 pp, $47 (P). [ISBN: 0-387-51299-3] Proceedings of the Fourth Bad Honnef Conference held in June 1988. Twenty-nine papers with thematic emphasis on adaptive methods in stochastic systems analysis and on the theory of random fields. JO

Partial Differential Equations, P. *Nonlinear Wave Equations.* Walter A. Strauss. CBMS Reg. Conf. Ser. in Math., No. 73. AMS, 1989, x + 91 pp, $21 (P). [ISBN: 0-8218-0725-0] A survey of and introduction to central aspects of nonlinear wave equations in absence of shocks. Chapters include invariance, singularities, solutions of small amplitude, scattering, stability of solitary waves, Yang-Mills equations, and Vlasov-Maxwell equations. SP

Partial Differential Equations, T(15: 1), L. *Introduction to Partial Differential Equations from Fourier Series to Boundary-value Problems.* Arne Broman. Dover, 1989, vi + 183 pp, $5.95 (P). [ISBN: 0-486-66158-X] An unabridged and corrected reproduction of the original (1968 Addison-Wesley text, TR, October 1971). Contains a bibliography new to this edition. MLR

Partial Differential Equations, P. *Lecture Notes in Mathematics-1402: Nonlinear Hyperbolic Problems.* Eds: C. Carasso, *et al.* Springer-Verlag, 1989, 249 pp, $24.30 (P). [ISBN: 0-387-51746-4] Proceed-

ings of a conference held in Bordeaux, France in June 1988. Contains sixteen papers (thirteen in English, three in French) addressing a variety of nonlinear problems, both numerically and analytically, ranging from fluid flow to stability of the Minkowski metric. SP

Partial Differential Equations, P. *Variational Methods for Boundary Value Problems for Systems of Elliptic Equations.* M.A. Lavrent'ev. Transl: J.R.M. Radok. Dover, 1989, 153 pp, $4.95 (P). [ISBN: 0-486-66170-9] Translation of a 1963 monograph, with corrections and an index. Approaches classical boundary value problems *via* geometric properties of conformal and quasi-conformal mappings, using the basic scheme of Hilbert and Tonnelli. For mathematicians and theoreticians in mechanics. DFA

Functional Analysis, T(16-17), S, L. *Functional Analysis with Applications.* B. Choudhary, Sudarsan Nanda. Wiley, 1989, xii + 344 pp, $34.95. [ISBN: 0-470-21564-X] Starts with introduction-level chapters on metric, normed and Hilbert spaces, Banach algebra and spectral theory, and operator theory. Concludes with chapters on fixed point theory, nonlinear operators, and sequence spaces. Intended as a text with exercises of varying difficulty closing each chapter. KS

Functional Analysis, T(15-17: 2), P, L. *Fourier Analysis.* T.W. Körner. Cambridge U Pr, 1988, xii + 591 pp, $34.50 (P). [ISBN: 0-521-38991-7] Paperback reprint of a 1988 text (TR, December 1988) containing over 100 short interconnected essays which, taken together, lead the reader through the theory and applications of Fourier analysis. The short chapters (the longest is ten pages) makes working through this text seem less foreboding. No formal exercises. SM

Analysis, P. *I.J. Schoenberg, Selected Papers.* Ed: Carl de Boor. Birkhauser Boston, 1988, $150 set [ISBN: 0-8176-3378-2]. *Volume 1,* xviii + 405 pp; *Volume 2,* xvi + 441 pp. This two-volume sampler contains about 45 of I.J. Schoenberg's 173 papers. Schoenberg's publications span the period 1928-1987; subjects range from splines to motion of billiard balls to placement of frets on guitars. Papers in *Volume 1* cover number theory, geometry, real analysis, and complex analysis. *Volume 2* emphasizes aspects of spline theory, especially cardinal splines. Interspersed in both volumes are brief commentaries by experts, summarizing Schoenberg's contributions in various areas. PZ

Analysis, S(18), P. *The Collected Works of Arne Beurling, Volume 2: Harmonic Analysis.* Eds: L. Carleson, *et al.* Birkhauser Boston, 1989, xx + 389 pp, $39. [ISBN: 0-8176-3412-6] The second of two volumes on complex analysis and harmonic analysis, respectively, of papers and seminar notes by Arne Beurling who did pioneering work in e.g., extremal length, inner functions, and spectral theory. The present volume contains 28 published papers, 1937-1970, together with hitherto unpublished seminar and lecture notes, the latter from a 1977-78 Mittag-Leffler series on harmonic analysis. A brief Foreword by L. Ahlfors and L. Carleson assesses Beurling's mathematical contributions. PZ

Differential Geometry, P. *Spin Geometry.* H. Blaine Lawson, Jr., Marie-Louise Michelsohn. Math. Ser., V. 38. Princeton U Pr, 1989, xii + 427 pp, $55. [ISBN: 0-691-08542-0] Presents the fundamental concepts of the geometry of spin manifolds (manifolds with a simply connected structure group) using Clifford algebras and their representations as a principle theme. The main tools used included the various index theorems of Atiyah and Singer and the theory of $C\ell_k$-linear elliptic operators. The last chapter is devoted to applications. MPR

Differential Geometry, T(18), S, P. *Lectures on Minimal Surfaces, Volume 1: Introduction, Fundamentals, Geometry and Basic Boundary Value Problems.* Johannes C.C. Nitsche. Cambridge University Pr, 1989, xxv + 563 pp, $135. [ISBN: 0-521-24427-7] An English translation from the original German text of the first five chapters of *Vorlesungen über Minimalflächen* supplemented with 66 pages of appendices and a collection of open problems. This broad survey of parametric minimal surfaces in Euclidean space is self-contained and includes considerable historical information. Bibliography contains nearly 1600 references. SP

Geometry, P. *Geometric Aspects of Banach Spaces.* Eds: E. Martin-Peinador, A. Rodés. London Math. Soc. Lect. Note Ser., V. 140. Cambridge U Pr, 1989, 194 pp, $24.95 (P). [ISBN: 0-521-36752-2] Contains 15 essays in honor of Antonio Plans. MLR

Geometry, T(16-17: 1, 2), L*. *Metric Affine Geometry.* Ernst Snapper, Robert J. Troyer. Dover, 1989, xx + 435 pp, $10.95 (P). [ISBN: 0-486-66108-3] Republication of the 1971 edition by Academic Press (TR, March 1972; Extended Review, March 1973). Based on Snapper course at First Cooperative Summer Seminar at Cornell University in 1964. Readers should have had a course in linear algebra and must know some elementary group theory. Exercises complement and extend the text. On the virtues of linear algebra as source for the axiom systems of all affine and projective geometries, both metric and nonmetric, the last three lines of the book say it best: "If he (Euclid) had lived today, what would he have done? He would have told us to do away with Hilbert's system and use linear algebra instead. And what a book he would have written!" JDEK

Geometry, P, L*. *Textual Studies in Ancient and Medieval Geometry.* Wilbur Richard Knorr. Birkhauser Boston, 1989, xvii + 852 pp, $119. [ISBN: 0-8176-3387-1] Complements the author's *The Ancient Tradition of Geometric Problems* (Birkhauser Boston, 1986, TR, October 1987). Exposition of the problems of cube duplication and angle trisection by Pappus and Eutocius. Treatment of Archimedes' *Dimension of the Circle* by Pappus and Theon and parallel versions in Arabic, Hebrew, and Latin. Three main parts: I. Ancient Texts on Geometric Problems; II. Arabic Geometric Texts and

Their Ancient Sources; and III. The Textual Tradition of Archimedes' *Dimension of the Circle*. Numerous notes and references. JDEK

Geometry, T*(13: 1). *Essentials of Geometry for College Students*. Margaret L. Lial, Arnold R. Steffensen, L. Murphy Johnson. Scott Foresman, 1990, xiii + 431 pp, $28. [ISBN: 0-673-38419-5] Geometry needed for further courses in college mathematics. Requires only beginning algebra. Easy to read; pace is slow. "Applications." Exercise sets, some of which are topped off with "old chestnuts"—a definite plus! Two-column proofs. Chapters on solid geometry, analytical geometry, and trigonometry. Nice collection of marginal notes and photographs. Good figures throughout. Test manual, solutions manual, transparencies are available for instructor; solution manual (odds plus) available for student. JDEK

Geometry, S(13-18), P, L.** *Recent Advances in Geometric Inequalities*. D.S. Mitrinović, J.E. Pečarić, V. Volenec. Math. & Its Applic. Kluwer Academic, 1989, xix + 710 pp, $149. [ISBN: 90-277-2565-9] The "Bible of Bottema" (*Geometric Inequalities*, O. Bottema, *et al.*, 1969; TR, February 1970; Extended Review, June-July 1971) contains about 400 varied geometric inequalities related to figures in the plane, and about 225 authors are cited. This monumental work updates and extends this beginning to several thousand inequalities, not only for figures in the plane but for figures in space and hyperspace as well, and over 750 names are cited. Moreover, the work contains extensive references, various methods of proof, as well as conjectures and unsolved problems. A voluminous and encyclopedic reference that should be in every mathematics library. LCL

Algebraic Topology, T(18: 1, 2), P. *Lecture Notes in Mathematics-1408: Transformation Groups and Algebraic K-Theory*. Wolfgang Lück. Springer-Verlag, 1989, xii + 443 pp, $42.70 (P). [ISBN: 0-387-51846-0] Self-contained introduction to the problem of classifying group actions on manifolds and its connection to algebraic K-theory. Divided into three main chapters: geometrically defined invariants, algebraically defined invariants, and $R\Gamma$-modules and geometry. Chapters 2 and 3 contain previously unpublished results. Each section begins with a brief summary of what is to be covered and ends with both an exercise set (containing some results not discussed in the section) and a comments section with references for further reading. SB

Topology, P. *Lecture Notes in Mathematics-1411: Topological Fixed Point Theory and Applications*. Ed: Boju Jiang. Springer-Verlag, 1989, vi + 203 pp, $20 (P). [ISBN: 0-387-51932-7] Proceedings of a conference at the Nankai Institute of Mathematics, Tianjin, China held April 1988. Twenty short papers from various areas of fixed point theory. SP

Dynamical Systems, P. *Dynamical Systems IV: Symplectic Geometry and its Applications*. Eds: V.I. Arnol'd, S.P. Novikov. Encycl. of Math. Sci., V. 4. Springer-Verlag, 1990, 283 pp, $59. [ISBN: 0-387-17003-0] Contains three somewhat independent sections translated from Russian. The first is a survey of the fundamental concepts of symplectic geometry. The latter two articles address applications to the theory of integrable systems and quantization. KS

Dynamical Systems, P. *The Connection Between Infinite Dimensional and Finite Dimensional Dynamical Systems*. Eds: Basil Nicolaenko, Ciprian Foias, Roger Temam. Contemp. Math., V. 99. AMS, 1989, xii + 357 pp, $41 (P). [ISBN: 0-8218-5105-5] Eighteen papers on the prospect that infinite-dimensional dynamics may not be too much harder than their finite-dimensional cousins (which are already hard enough). BC

Dynamical Systems, S(16-17), L. *Statistical Analysis and Control of Dynamic Systems*. Hirotugu Akaike, Toichiro Nakagawa. Trans: Hirotugu Akaike, Maki Akaike Momma. Math. & Its Applic. Kluwer Academic, 1988, xiii + 210 pp, $99. [ISBN: 90-277-2786-4] Nicely illustrates use of statistical techniques to automating operation of cement kilns. Carefully defines problem and develops necessary statistical background. Includes Fortran listing and review of the TIMSAC package. Note price. RWJ

Control Theory, S(18), P. *Adaptive Control & Identification: Conflict or Confluz?* J.W. Polderman. CWI Tract, V. 67. Mathematisch Centrum, 1989, 117 pp, Dfl. 17,80 (P). [ISBN: 90-6196-377-X] Concerned with problems of adaptive control. What distinguishes this theory from the classical control theory is that the laws governing the systems to be controlled are no longer completely known. (Reprint of a doctoral thesis.) KS

Control Theory, T(15-17), L. *Linear Systems and Optimal Control*. C.K. Chui, G. Chen. Ser. in Inform. Sci., V. 18. Springer-Verlag, 1989, viii + 155 pp, $59.50. [ISBN: 0-387-18737-5] Seeks to present a brief and reasonably complete study of linear systems in both continuous time and discrete time settings. Optimal control of these systems is also covered. The usual topics are covered, namely state space descriptions, controllability, observability, stability and optimal control via variational techniques. Plenty of exercises after each chapter. SM

Control Theory, P, L. *Lecture Notes in Control and Information Sciences-136: Stochastic Systems and Optimization*. Ed: J. Zabczyk. Springer-Verlag, 1989, vi + 373 pp, $59.50 (P). [ISBN: 0-387-51619-0] Contains the proceedings of the Sixth IFIP Working Conference on Stochastic Systems and Optimization which was held in Warsaw, Poland during September 1988. The papers are organized into three categories: finite-dimensional systems, infinite-dimensional systems, and control and estimation. SM

Stochastic Processes, P. *Optimization of Stochastic Systems: Topics in Discrete-Time Dynamics, Second Edition*. Masanao Aoki. Economic Theory, Econometrics, & Math. Econ. Academic Pr, 1989, x + 417 pp, $64.95. [ISBN: 0-12-058851-X] A graduate-level introduction to sequential decision

analysis under uncertainty. This edition has been extensively revised to make it more accessible to graduate students and professionals in the social sciences (especially economics). MLR

Elementary Statistics, T(13-14: 1, 2). *Introduction to Statistics.* Anthony A. Salvia. Saunders College, 1990, xv + 603 pp, $45.25. [ISBN: 0-03-012634-7] Covers descriptive statistics, basic probability models, estimation, hypothesis testing (including ANOVA and chi-squared tests), correlation, and regression. Chapters generally contain a section devoted to nonparametric procedures and indicate when these may be preferred to the corresponding parametric procedures. Some computer output including MINITAB. Requires only high school algebra. RWJ

Statistics, S(18), P. *Bayesian Analysis of Time Series and Dynamic Models.* Ed: James C. Spall. Statistics: Textbooks & Monographs, V. 94. Marcel Dekker, 1988, xxvii + 536 pp, $99.75. [ISBN: 0-8247-7936-3] This book presents Bayesian techniques for analyzing dynamic systems, systems which change over time. Each chapter is written by a different author or authors who are experts in the fields of statistics, control systems, engineering, or econometrics. The models presented can be broken up into two general types: traditional ARMA time series models, and linear and nonlinear state-space models. The book begins with an historical overview of the field. Each chapter contains references to other work in the field. JJ

Programming, P. *Ada Quality and Style: Guidelines for Professional Programmers.* The Software Productivity Consortium. Van Nostrand Reinhold, 1989, ix + 230 pp, $26.95 (P). [ISBN: 0-442-23805-3] A well-focused technical essay on such things as indentation, notational conventions, and programming style with the aim of promoting readable, reliable, and portable Ada programs. JAS

Programming, P, L. *Weaving a Program: Literate Programming in WEB.* Wayne Sewell. Van Nostrand Reinhold, 1989, xx + 556 pp, $34.95 (P). [ISBN: 0-442-31946-0] A comprehensive overview of the WEB programming environment created by Donald Knuth as a tool for literate programming. A WEB source file is a mixture of program code fragments (usually in Pascal) and explanatory text (usually in TEX) that is processed by a program called "weave" to produce a TEX input file which will yield neatly printed human-readable listings of the WEB source file to serve as documentation for the program, which is itself compiled from Pascal source code (not human-readable) that is created from the *same* WEB source file by the program "tangle." Discusses issues of implementation on various machines and of adaptations to different languages. Includes source listings for "weave" and "tangle," and many useful examples, tables, and details. LAS

Languages, P. *SGML: An Author's Guide to the Standard Generalized Markup Language.* Martin Bryan. Addison-Wesley, 1988, xvii + 364 pp, (P). [ISBN: 0-201-17535-5] As advertized, a description and explanation of this ISO standard for a computer description of documents for the publishing industry as adopted by the U.S. Department of Defense and European Community Office of Official Publications. JAS

Computer Systems, T(14-15: 1), L*. *Computation Structures.* Stephen A. Ward, Robert H. Halstead, Jr. MIT Pr, 1990, xx + 789 pp, $49.95. [ISBN: 0-262-23139-5] This textbook was developed for use in the sophomore "core" course taken by all computer science and electrical engineering undergraduates at M.I.T. Covers the design and implementation of computers from the level of logic gates to machine language. A very interesting book. AO

Computer Graphics, T(17-18: 1, 2), P. *The Mathematical Structure of Raster Graphics.* Eugene L. Fiume. Academic Pr, 1989, xvi + 221 pp, $29.95. [ISBN: 0-12-257960-7] A step, possibly a big one, towards a formal mathematical system for raster graphics. Requires a knowledge of mathematics through the beginning graduate level in algebra and analysis as well as upper-level experience with graphics including illumination. JAS

Artificial Intelligence, P. *Applications of Expert Systems.* Ed: J. Ross Quinlan. Addison-Wesley, 1987, xvi + 223 pp. [ISBN: 0-201-17449-9] Papers from the Second Australian Conference on Applications of Expert Systems. The focus is on the move from the experimental expert system to the nature and role of the commercial tool. JAS

Artificial Intelligence, P. *Case-Based Planning: Viewing Planning as a Memory Task.* Kristian J. Hammond. Perspect. in Artif. Intellig., V. 1. Academic Pr, 1989, xxiv + 277 pp, $34.95. [ISBN: 0-12-322060-2] Case-based planning is based on the idea of modifying an old plan rather than building a new one from first principles. Memory of past successes is used to find the best plan to modify for current goals, and problems and new successes are saved and indexed so they can be used to satisfy similar goals and avoid similar problems. Theory is implemented in the program CHEF, whose domain is Szechuan cooking. RBK

Computer Science, P. *Formal Development of Programs and Proofs.* Ed: Edsger W. Dijkstra. Addison-Wesley, 1990, xiv + 241 pp. [ISBN: 0-201-17237-2] Papers from, inspired by, and related to the sixth (of six) Programming Institute which was part of the year of Programming at the University of Texas at Austin. The focus is on those aspects of programming which appear to be mathematically treatable. JAS

Computer Science, P. *Lecture Notes in Computer Science-393: Categorical Methods in Computer Science With Aspects from Topology.* Eds: H. Ehrig, *et al.* Springer-Verlag, 1989, vi + 350 pp, $28.10 (P). [ISBN: 0-387-51722-7] Proceedings of a September 1988 international workshop held in (West) Berlin. JAS

Applications, S(18), P. *Mathematical Psychology*

in Progress. Ed: Edward E. Roskam. Recent Res. in Psych. Springer-Verlag, 1989, viii + 385 pp, $53 (P). [ISBN: 0-387-51686-7] Third volume of papers originating from recent annual meetings of the European Mathematical Psychology Group. Papers partitioned into three groups: theory, metatheory, and measurement; choice, perception, cognition, and performance; psychometrics and theory of data. Papers represent "work in progress" and are high in mathematical content. WE

Applications, P, L. *Mathematics in Industrial Problems, Part 2.* Avner Friedman. IMA, V. 24. Springer-Verlag, 1989, xii + 183 pp, $24. [ISBN: 0-387-97139-4] A collection of eighteen mathematical problems motivated by specific "grass roots" situations encountered in industry. A very wide range of problems are encountered, the common thread being the application of mathematical method to real problems as opposed to the imagined problems often encountered in textbooks. Each of the eighteen chapters is self-contained. The collection ends with brief discussions of solutions to problems raised in the first volume of this work. SM

Applications, P. *Lecture Notes in Statistics-61: The Analysis of Directional Time Series: Applications to Wind Speed and Direction.* Jens Breckling. Springer-Verlag, 1989, viii + 238 pp, $23.80 (P). [ISBN: 0-387-97182-3] The first half of the book concentrates on the empirical aspects of directional time series by performing a comprehensive analysis of hourly records of wind speed and direction at the port of Fremantle, Australia. The wind records can be explained by a geostrophic component and daily component, both related to the season and weather. The second half of the book looks at the problem in general from a theoretical standpoint using two types of models: the von Mises and wrapped (normal) autoregressive processes. The monograph is based on the author's Ph.D. thesis. JJ

Applications, P. *Theory and Practice of Geometric Modeling.* Eds: Wolfgang Straßer, Hans-Peter Seidel. Springer-Verlag, 1989, x + 547 pp, $89.50. [ISBN: 0-387-51472-4] Expanded lectures from a conference held in Blaubeuren, West Germany in October 1988. Emphasis on bridging the gap "between theoretical results, performance of existing [Computer Aided Design] systems and the real problems of users" leads to subjects ranging from abstract differential geometry and analysis of algorithms to descriptions of new software. Some nice four color graphics. JO

Applications (Biological Science), P. *Lecture Notes in Biomathematics-80: Depth Perception in Frogs and Toads: A Study in Neural Computing.* Donald House. Springer-Verlag, 1989, vii + 135 pp, $16.20 (P). [ISBN: 0-387-97157-2] Models for monocular and binocular cooperation in neural computing algorithms for navigation and prey localization, with implications for animal experiments and robotics. BC

Applications (Biological Science), P, L. *Lecture Notes in Statistics-55: Estimation and Analysis of Insect Populations.* Eds: L. McDonald, *et al.* Springer-Verlag, 1989, xiv + 492 pp, $54 (P). [ISBN: 0-387-96998-5] Consists of thirty-five papers presented at the symposium/workshop on The Estimation and Analysis of Insect Population held in January 1988 in Laramie, Wyoming. The papers evaluate current methodological and present promising new approaches. Four subject areas are represented: analysis of stage-frequency data, modeling of population dynamics, analysis of spatial data, and general sampling and estimation methods. SM

Applications (Fluid Dynamics), P. *Annual Review of Fluid Mechanics, Volume 22, 1990.* Eds: John L. Lumley, Milton van Dyke, Helen L. Reed. Annual Reviews, 1990, x + 592 pp, $38. [ISBN: 0-8243-0722-4] Essay on the history of the Reynolds number, followed by seventeen recent papers on topics including the design of yachts, human-powered aircraft, and offshore structures. JO

Applications (Information Theory), P. *Lecture Notes in Control and Information Sciences-129: Advances in Communications and Signal Processing.* Eds: W.A. Porter, S.C. Kak. Springer-Verlag, 1989, vi + 376 pp, $47 (P). [ISBN: 0-387-51424-4] Proceedings of an international conference held in Baton Rouge, Louisiana during October 1988. Contains thirty-four selected papers written by researchers in various fields. Other papers appear in *Advances in Computing and Control* (TR, February 1990), also in the Springer-Verlag *LNCIS* series. SP

Applications (Social Science), S(16), C, P. *Simulating K=3 Christaller Central Place Structures: An Algorithm Using a Constant Elasticity of Substitution Consumption Function.* Daniel A. Griffith. Mono., No. 10. Institute of Mathematical Geography (2790 Briarcliff, Ann Arbor, MI 48105), 1989, v + 103 pp, $15.95 (P). This monograph presents a description of and the source code for a program that simulates K=3 Christaller central place structures using a constant elasticity of substitution consumption function. It includes a discussion of the salient geographic theory and mathematical underpinnings. Also includes a list of references. The source code is in Basic for PC compatibles. CEC

Reviewers

DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; WE: William Etter, Macalester; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; RH: Rhonda Hatcher, St. Olaf; RWJ: Roger W. Johnson, Carleton; JJ: Jason Jones, St. Olaf; RBK: Roger B. Kirchner, Carleton; JDEK: Joseph D.E. Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; JO: Jeff Ondich, St. Olaf; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; FR: Flauren Ricketts, Macalester; KS: Karen Saxe, St. Olaf; JS: John Schue, Macalester; JAS: J. Arthur Seebach, Jr., St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.



Carus Mathematical Monograph No. 20

THE GENERALIZED RIEMANN INTEGRAL

Robert M. McLeod
275 + xiii pages. Hardbound.
List: $24.95 MAA Member: $22.00

Until recently the most powerful and beautiful tools of integration theory have been accessible only to the privileged few whose studies extended through the Lebesgue integral. Now a new integral, a generalization of the familiar Riemann integral, has been discovered which has all the power and range of the Lebesgue integral but which can be readily understood by anyone who has studied calculus through the sophomore level.

The Generalized Riemann Integral is the first book-length presentation of this exciting new development in integration theory. Because of the clarity and organization of the exposition and the inclusion of exercises designed to actively engage the reader in the material, it is eminently suitable for use as a textbook at the advanced undergraduate or beginning graduate level. Furthermore, because it presents within a single volume results which were previously scattered throughout many research publications, it will undoubtedly also be of considerable interest to specialists in integration theory.

List of Contents: Definition of the Generalized Riemann Integral • Basic Properties of the Integral • Absolute Integrability and Convergence Theorems • Integration of Subsets of Intervals • Measurable Functions • Multiple and Iterated Integrals • Integrals of Stieltjes Type • Comparison of Integrals • Appendix: Solutions of In-Text Exercises

THE MATHEMATICAL ASSOCIATION OF AMERICA

Order From:
Mathematical Association of America
1529 Eighteenth Street NW
Washington DC 20036



Complex Variables

Complex plane or Riemann sphere.

See how functions combine - apply composition, integration, differentiation - in any combination.

***f(z)* can handle poles and knows how to pick a continuous branch of the logarithm.**

Generate animation as t paramter varies.

Watch the actual complex "graph" as it rotates in *four* dimensions!

Includes *f(z)*-Mandelbrodt

From our mail...

"Your f(z) software is one of the most interesting mathematical tools I have come across...It's a whole new mathematical world." S. S. M.

"Excellent, beyond my imaginings...many thanks once again for reopening my eyes to the many beauties of mathematics. . . I had never dreamed it would be so readily and easily available." D. H. - Australia

"f(z) is spectacular..I only wish I had had this when I was back in college!!!" V. W. - Raychem

"Great program, terrific graphics, never seen anything quite like it before." S. P.

For PC or Macintosh

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

TOPOLOGY OF 4-MANIFOLDS

Michael Freedman and Frank Quinn

One of the great achievements of contemporary mathematics is the new understanding of four dimensions. Michael Freedman and Frank Quinn have been the principals in the geometric and topological development of this subject, proving the Poincaré and Annulus conjectures respectively. In *Topology of 4-Manifolds* these authors have collaborated to give a complete and accessible account of the current state of knowledge in this field.

Princeton Mathematical Series, 39. 79 illustrations.

Cloth: $49.50 ISBN 0-691-08577-3

SPIN GEOMETRY

H. Blaine Lawson, Jr., and Marie-Louise Michelsohn

This book offers a systematic and comprehensive presentation of the concepts of a spin manifold, spinor fields, Dirac operators, and -genera, which, over the last two decades, have come to play a significant role in many areas of modern mathematics. Since the deeper applications of these ideas require various general forms of the Atiyah-Singer Index Theorem, the theorems and their proofs, together with all prerequisite material, are examined here in detail.

Princeton Mathematical Series, 38

Cloth: $55.00 ISBN 0-691-08542-0



Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900

ORDERS 800-PRS-ISBN (777-4726)

GEOMETRY & TOPOLOGY

FIRST CONCEPTS OF TOPOLOGY: The Geometry of Mappings of Segments, Curves, Circles, and Disks

W.G. Chinn and N.E. Steenrod

This clear and winning little book, for readers willing to come to genuine grips with the idea of a mathematical proof, presents topology . . . as mathematicians see it. The parlor tricks are gracefully alluded to here and there, but they are distinctly for after hours. The center of interest is the stuff itself: powerful notions of set theory are . . . exploited to define open sets and their coverings, and from them [to prove] the key theorems. .. One cannot any longer doubt that a single stroke of a knife exists that divides any irregular ham sandwich so that ham and both bread slices can be shared with perfect fairness by two consumers.

Philip Morrison in *Scientific American*

Contains over one hundred and fifty problems and solutions.

160 pp., 1966, ISBN 0-88385-618-2

List: $10.00 MAA Member: $8.00

Catalog Number NML-18

FROM PYTHAGORAS TO EINSTEIN

K.O. Friedrichs

Starting with area and dissection proofs of the Pythagorean theorem, Friedrichs proceeds gently, step by step, to cover vectors, coordinates, elastic and inelastic impacts, and relativistic space-time–ending with a derivation of a contemporary formula rivaling the Pythagorean theorem in fame, $E = mc^2$.

88 pp., 1965, ISBN 0-88385-616-6

List: $9.50 MAA Member: $7.60

Catalog Number NML-16

GEOMETRY REVISITED

H.S.M. Coxeter and S.L. Greitzer

After seeing a proof that a segment has a unique midpoint, students may wonder whether there are any interesting theorems in geometry. If either they or their teachers get a hold of this book, they will learn that there are many beautiful and nontrivial theorems in geometry. Among those found here are the theorems of Ceva, Menelaus, Pappus, Desargues, Pascal, and Brianchon. A nice proof is given of Morley's remarkable theorem on angle trisectors. The transformational point of view is emphasized: reflections, rotations, translations, similarities, inversions, and affine and projective transformations. Many fascinating properties of circles, triangles, quadrilaterals, and conics are developed.

193 pp., 1967, ISBN 0-88385-619-0

List: $11.75 MAA Member: $9.50

Catalog Number NML-19

The very lucid presentation takes the reader from the elementary problems of plane Euclidean geometry to the fundamental concepts of non-Euclidean geometry . . . The book is rich in remarkable facts and thereby is very effective in promoting the significance and the value of geometry in mathematical teaching, a promotion which is very necessary . . . The always original developments use very simple tools . . . and soon proceed to higher configurations.

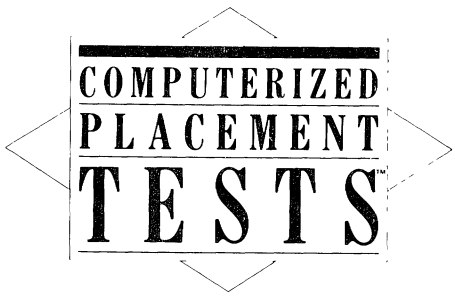
K. Strubecker in *Mathematical Reviews*

Order from:



The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036
(202) 387-5200

NEW! Accurate advanced math placement In as few as 20 minutes!



COLLEGE-LEVEL MATHEMATICS

This fall you can place your entry-level students in advanced mathematics courses in as few as 20 minutes per student! There's no quicker, easier, or more precise method to find out what they've mastered and whether they're ready for intermediate algebra, college algebra, pre-calculus, or first-level calculus.

Now—preview this exciting breakthrough in the College Board's Computerized Placement Tests!

Here is your opportunity to preview a set of examination disks—and experience a powerful innovation in computer adaptive testing called *Seamless Serial Testing*. It tests students across a wide range of concepts with fewer questions than conventional tests, and provides more accurate measurement than conventional tests. Based on their answers, individual students are routed automatically to more advanced or more basic material. You can administer this new test module independently or combine it with lower-level CPTs in Arithmetic and Elementary Algebra. Operates on IBM-PC, XT, AT, PS/2 and compatibles, including portables and laptops.

Order your set of examination disks today!

See for yourself how *College-Level Mathematics* and the other four CPTs in Reading, Writing, Arithmetic, and Elementary Algebra can handle all your institution's placement needs in a fraction of the time it now takes. Your examination set is just $25 (credited toward your first order). You'll also receive a FREE diskette demonstrating *Accuplacer*, the College Board's exclusive computerized assessment, placement, and management software. Please specify 5 1/4" or 3 1/2" disks.

The College Board
Attn: L.J. Abernathy
45 Columbus Avenue
New York, New York 10023-6992

Copyright © 1990 by the College Entrance Examination Board. College Board and the acorn logo are registered trademarks of the College Entrance Examination Board. Accuplacer, College Board Computerized Placement Test, and Seamless Serial Testing are trademarks owned by the College Entrance Examination Board.

INTERNATIONAL MATHEMATICAL OLYMPIADS

The International Mathematical Olympiads, 1959-1977 and 1978-1985 provide a compilation of problems and solutions of arresting ingenuity; all accessible to secondary school students. The alternative solutions are particularly interesting because they show that there are many ways to solve a problem.

INTERNATIONAL MATHEMATICAL OLYMPIADS, 1959-1977

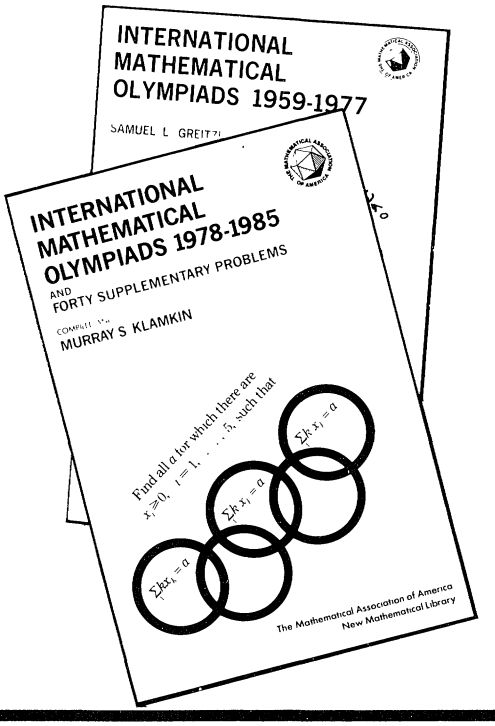
Compilation and solutions by S.L. Greitzer

116 problems

204 pp., 1978, ISBN 0-88385-627-1

List: $11.50 MAA Member: $9.50

Catalog Number NML-27



INTERNATIONAL MATHEMATICAL OLYMPIADS , AND FORTY SUPPLEMENTARY PROBLEMS, 1978-1985

Compilation, solutions, and 40 additional problems by Murray S. Klamkin

88 problems in all

150 pp., 1986, ISBN 0-88385-631-X

List: $12.95 MAA Member: $10.50

Catalog Number NML-31

This sequel of problems used in the International Mathematical Olympiad certainly matches the description 'meant to challenge.' The problems contained in this publication are the most challenging I have seen. For the serious problem solver, this is an excellent collection of truly challenging problems designed to give instruction in the use of sophisticated methods and modes of attack.

Richard J. Wylie in *The Mathematics Teacher*

Both International Olympiad books available as a package:

List: $21.00 MAA Member: $17.50

Catalog Number IMO

Order from:
The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

Math & More

GARDNER'S WHYS AND WHEREFORES
Martin Gardner

The well-known wit and encyclopedic knowledge of Martin Gardner are on display as he tackles topics from the "anthropic principle" to the idiosyncrasies of Allan Bloom, critiques fantasy writers from Lord Dunsany to Ray Bradbury, and questions whether physicists are really on the verge of discovering Everything.

Cloth $19.95 272 pages

Now in Paper—

MATHEMATICS AND THE UNEXPECTED
Ivar Ekeland

With a Foreword by Felix Browder

"Highly engaging and entertaining. . . . Ekeland uses celestial mechanics, Poincaré maps for the Henon system, the cat map, the Lorenz model, and Sitnikov's three-body problem all to good effect."—Ronald F. Fox, *American Journal of Physics*

Paper $8.95 160 pages 5 halftones, 39 line drawings

MATHEMATICS AND HUMOR
John Allen Paulos

"[An] illuminating account of the relationship between humor and mathematics."—Joseph Williams, *New York Times*

Paper $10.95 124 pages 6 line drawings, 29 figures

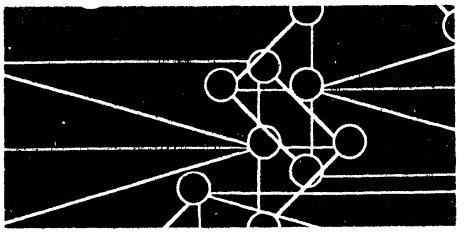
University of Chicago Press
5801 South Ellis Avenue Chicago, IL 60637

Math Texts

ESSENTIAL RESULTS OF FUNCTIONAL ANALYSIS
Robert J. Zimmer

This slim book concisely presents fundamental ideas of functional analysis, providing essential results from the many domains of mathematics and mathematical physics to which functional analysis makes contributions. Zimmer summarizes measure theory and the elementary theory of Banach and Hilbert spaces, then discusses topological vector spaces, seminorms defining them, and natural classes of linear operators. He then presents basic results for a wide range of topics: convexity and fixed point theorems, compact operators, compact groups and their representations, spectral theory of bounded operators, ergodic theory, commutative C*- algebras, Fourier transforms, Sobolev embedding theorems, distributions, and elliptic differential operators.

Paper $14.95 (est.) 168 pages (est.)
Library cloth edition $34.95 (est.)
Chicago Lectures in Mathematics series



Now in Paper

THE THEORY OF LINEAR ECONOMIC MODELS
David Gale

This outstanding introduction to mathematical economics is now available in a paperback edition.

"An excellent introduction to linear programming, Neumann models, and matrix games, as well as to the theory of linear inequalities which is fundamental for these subjects."—Lionel W. McKenzie, University of Rochester

Paper $14.95 356 pages

University of Chicago Press
5801 South Ellis Avenue Chicago, IL 60637



Random Walks and Electric Networks,

by J. Laurie Snell and Peter Doyle

xiii + 159 pages. Hardbound
List: $25.00 MAA Member: $19.00

In this newest addition to the Carus Mathematical Monographs, the authors examine the relationship between elementary electric network theory and random walks, at a level which can be appreciated by the able college student. We are indebted to them for presenting this interplay between probability theory and physics in so readable and concise a fashion.

Central to the book is Polya's beautiful theorem that a random walker on an infinite street network in d-dimensional space is bound to return to the starting point when d = 2, but has a positive probability of escaping to infinity without returning to the starting point when d = 3. The authors interpret this theorem as a statement about electric networks, and then prove the theorem using techniques from classical electrical theory. The techniques referred to go back to Lord Rayleigh who introduced them in connection with an investigation of musical instruments.

In Part I the authors restrict themselves to the study of random walks on finite networks, establishing the connection between the electrical concepts of current and voltage and corresponding descriptive quantities of random walks regarded as finite state Markov chains. Part II deals with the idea of random walks on infinite networks.

Table of Contents

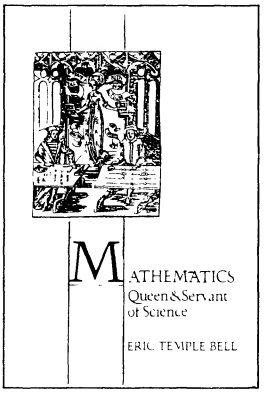
***Part I*: Random Walks on Finite Networks**
Random Walks in One Dimension
Random Walks in Two Dimensions
Random Walks on More General Networks
Rayleigh's Monotonicity Law

***Part II*: Random Walks on Infinite Networks**
Pólya's Recurrence Problem
Rayleigh's Short-Cut Method
The Classical Proofs of Polya's Theorem
Random Walks on More General Infinite Networks



Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

MATHEMATICS & BIOGRAPHY



MATHEMATICS: QUEEN AND SERVANT OF SCIENCE

E.T. Bell

An absorbing account of pure and applied mathematics from the geometry of Euclid to that of Riemann and its application in Einstein's theory of relativity. The twenty chapters treat such topics as: algebra, number theory, logic, probability, infinite sets and the foundations of mathematics, rings, matrices, transformations, groups, geometry, and topology. Republished in 1987 with corrections and an added Foreward by Martin Gardner.

454 pp., ISBN 0-88385-446-3
Paperbound

List: $15.95 MAA Member: $11.95

Catalog Number QAS

The book deserves a place in today's market. It is a much more popular work than most histories of the subject, and that is exactly what makes it accessible to students as well as to non-mathematicians. It is rewarding reading for . . . teachers and students at all mathematical levels.

Morris Kline of The Courant Institute

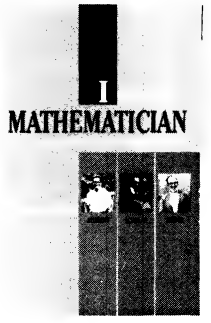
I WANT TO BE A MATHEMATICIAN
An Automathography in Three Parts

Paul R. Halmos

This is a book to be read with interest by all those who know, or might want to know, what mathematicians and mathematical careers are like. Paul R. Halmos begins with his school days and carries the reader swiftly through a career that has sustained itself at a high level since his first post-doctoral days at the Institute for Advanced Study in 1939, where he worked with John von Neumann among others. Still going strong in 1988, Halmos has contributed much to logic, operator theory, ergodic theory, and the literature in general.

442 pp., 1988, Paper, ISBN 0-88385-445-7

List: $18.00 MAA Member: $15.00



Catalog Number IWM

It is a truly unique book, which nobody but Paul Halmos could have written. I think it will be a classic.

Constance Reid

The book is exciting, witty, and well worth the time invested in its study. It communicates what it means to be a mathematician.

John Dossey in *The Mathematics Teacher*

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036
(202) 387-5200

Studies in the History of Mathematics, Esther R. Phillips, Editor.
Volume #26 in MAA Studies in Mathematics
320 pp., Hardbound, ISBN-0-88385-128-8.
Catalog Number - MAS-26
List: $36.50 MAA Member: $28.00

Esther R. Phillips has brought together a collection of articles showing the sweep of recent scholarship in the history of mathematics. The material covers a wide range of current research topics: algebraic number theory, geometry, topology, logic, the relationship between mathematics and computing, partial differential equations, and algebraic geometry. This volume will show teachers and their students the historical background and the development of advanced topics in mathematics and of the major fields of contemporary research. Each article contains references for further study.

Dedekind's Invention of Ideals Harold M. Edwards

Non-Euclidean Geometry and Weierstrassian Mathematics: The Background of Killing's work on Lie Algebras Thomas Hawkins

The Discovery of Non-Euclidean Geometry Jeremy Gray

L.E.J. Brouwer's Coming of Age as a Topologist Dale M. Johnson

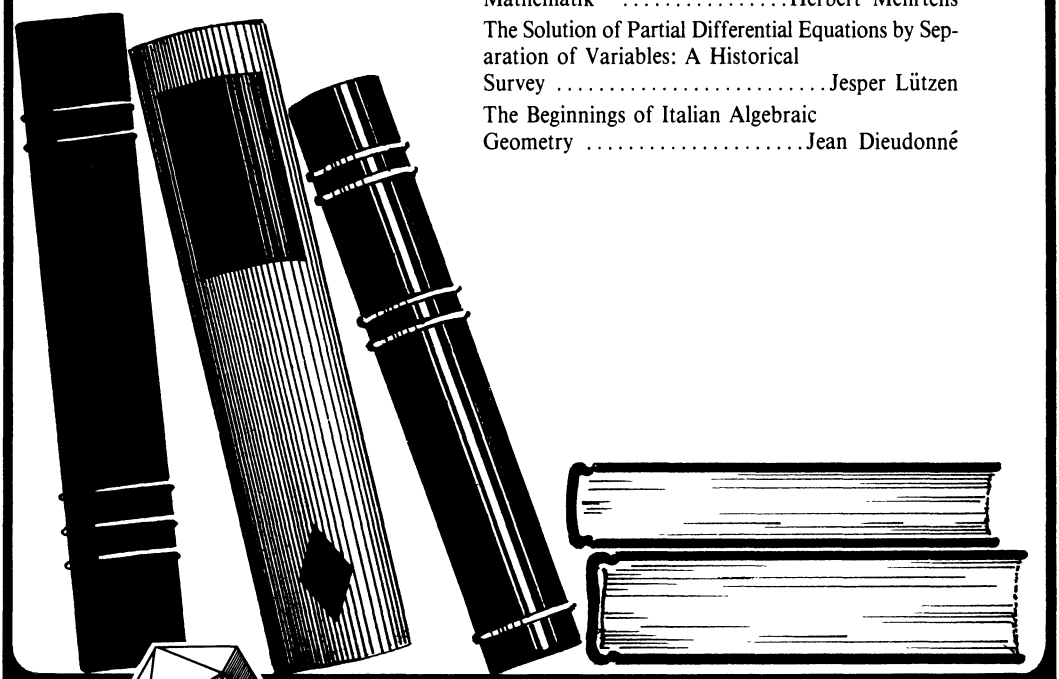
A House Divided Against Itself: The Emergence of First-Order Logic as the Basis for Mathematics Gregory H. Moore

The Mathematical Reception of the Modern Computer: John von Neumann and the Institute for Advanced Study Computer William Aspray

Ludwig Bieberbach and "Deutsche Mathematik" Herbert Mehrtens

The Solution of Partial Differential Equations by Separation of Variables: A Historical Survey Jesper Lützen

The Beginnings of Italian Algebraic Geometry Jean Dieudonné



Order from
The Mathematical Association of America
1529 Eighteenth St. NW
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 6 June–July 1990

Contents

(ISSN 0002-9890)

ARTICLES

The Geometry of Harmonic Oscillators KENNETH R. MEYER 457

Game of Cards, Dynamical Systems, and a Characterization of the Floor and Ceiling Functions P. EISELE AND K. P. HADELER 466

Decomposition Theory in the Teaching of Elementary Linear Algebra R. R. LONDON AND H. P. ROGOSINSKI 478

Representing Binomial Coefficients as Sums of Squares ANDREW GRANVILLE AND YILIANG ZHU 486

LETTERS TO THE EDITOR 494

NOTES

A Geometric Regularity From Order-Restricted Statistical Inference TIM ROBERTSON AND CAROLYN PILLERS 496

The Solution of Certain Integral Equations by Means of Operators of Arbitrary Order BERTRAM ROSS AND BALDEV K. SACHDEVA 498

Relations Between $\omega(n)$ and $\Omega(n)$ ARMEL MERCIER 503

Gear Trains and Continued Fractions JEROLD MATHEWS 505

THE TEACHING OF MATHEMATICS

Overdetermined Systems of Linear Equations GARETH WILLIAMS 511

When Quasinormal Implies Normal DEAN HICKERSON, SHERMAN STEIN, AND KENYA YAMAOKA 514

A Geometric Approach to l'Hôpital's Rule GIANLUCA GORNI 518

Understanding Vector Fields C. R. CURJEL 524

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 528

Advanced Problems and Solutions 535

REVIEWS

Perspectives of Nonlinear Dynamics. By E. Atlee Jackson JOHN M. GUCKENHEIMER 544

TELEGRAPHIC REVIEWS 547

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

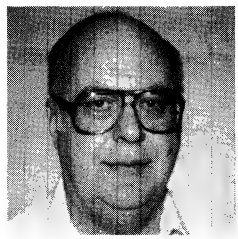
Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

The Geometry of Harmonic Oscillators†

KENNETH R. MEYER, *University of Cincinnati*



KENNETH R. MEYER: I received my Ph.D. from the University of Cincinnati in 1964, was an Assistant Professor at Brown University, was an Associate Professor at the University of Minnesota and am Professor at the University of Cincinnati. I learned the geometry of harmonic oscillators from the late Charlie Conley.

I. Introduction. About a year ago an engineering colleague dropped by my office to ask some questions about the analysis of a system of differential equations. After hearing the problem, I said he needed to know about the geometry of pairs of harmonic oscillators. He looked a little hurt at my suggestion, but consented to indulge me for a while. After I lectured at the chalk board for about an hour, he admitted that before my lecture he had thought he understood the harmonic oscillator, but now it was clear that he hadn't.

Although an introduction to the harmonic oscillator comes early in the study of the calculus and it is extensively used throughout physics, the beautiful geometry of harmonic oscillators is seldom given in differential equations texts of any level. Even some researchers in oscillation theory are unaware of the interrelationship between the Hopf fibration, the ergodic flow on a torus, the rational approximations of irrational numbers, the topology of the 3-sphere and the lowly harmonic oscillator. The material presented below has been known since the time of Poincaré and Birkhoff and so nothing given here is new. It is essential in Conley's analysis of periodic solutions of the restricted three body problem [**4**], Weinstein's normal mode theorem [**11**] and the generalizations of these theorems found in [**5**, **7**, **8**].

First I will present the geometry of a single harmonic oscillator from a dynamical systems point of view with particular emphasis on the use and interpretation of polar coordinates. Then a model for the 3-sphere as an identification space is obtained by using polar coordinates for a pair of harmonic oscillators. Many basic geometric facts can be gleaned from this simple model of the 3-sphere. The Hopf fibration, the knotting and linking of solutions, and the irrational flow on the torus are a few of the topics presented. The classical regularization of the Kepler problem (the central force problem) is given as an example of a non-trivial physical problem which can be reduced to a pair of harmonic oscillators. Thus the geometry discussed here sheds some light on the classical two body problem.

II. One Harmonic Oscillator. From the point of view of dynamical systems theory, a differential equation defines a flow on a space and the solutions are curves in that space. Historically, the problems come from classical dynamics and so the independent variable is 'time' and will be denoted by t. Newton's notation for fluxions persists with

$$\dot{} = \frac{d}{dt} \quad \text{and} \quad \ddot{} = \frac{d^2}{dt^2}. \tag{1}$$

This work was supported by a grant from ACMP/DARPA administered by NBS.

Thus, the harmonic oscillator,

$$\ddot{x} + \omega^2 x = 0, \tag{2}$$

where ω is a positive constant, is written as a system of first order equations by introducing the new variable $u = \dot{x}/\omega$, to become

$$\begin{aligned} \dot{x} &= \omega u \\ \dot{u} &= -\omega x. \end{aligned} \tag{3}$$

The variable u is a scaled velocity and so the x, u plane is essentially the position-velocity plane, or the phase space of physics. The basic existence and uniqueness theorem of differential equations asserts that through each point (x_0, u_0) in the plane there is a unique solution which passes through this point at any particular epoch t_0. These solutions are given by the formula

$$\begin{pmatrix} x(t, t_0, x_0, u_0) \\ u(t, t_0, x_0, u_0) \end{pmatrix} = \begin{pmatrix} \cos \omega(t - t_0) & \sin \omega(t - t_0) \\ -\sin \omega(t - t_0) & \cos \omega(t - t_0) \end{pmatrix} \begin{pmatrix} x_0 \\ u_0 \end{pmatrix}. \tag{4}$$

The solution curves are parameterized circles. The reason that one introduces the scaled velocity instead of using the velocity itself, as is usually done, is so that the solution curves become circles instead of ellipses. In dynamical systems the geometry of this family of curves in the plane is of prime importance.

The system admits an integral (energy in physical problems)

$$I = x^2 + u^2 \tag{5}$$

which is constant along the solutions of (3), since $\dot{I} = 2x\dot{x} + 2u\dot{u} = 0$ by virtue of (3). Since a solution lies in the set where $I =$ constant, a circle in the x, u plane, the integral alone gives the geometry of the solution curves in the plane. See FIGURE 1.

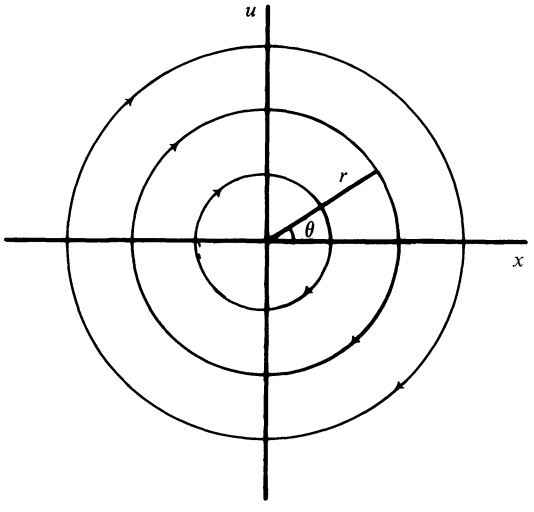


FIG. 1

If we introduce polar coordinates $r^2 = x^2 + u^2$, $\theta = \tan^{-1} u/x$ then the equations (3) become

$$\dot{r} = 0, \qquad \dot{\theta} = -\omega. \tag{6}$$

Here again we see that the solutions lie on circles about the origin since, $\dot{r} = 0$, and that the circles are swept out with constant angular velocity.

In many cases the harmonic oscillator is only a first approximation to a physical problem. For example one gets a harmonic oscillator from the pendulum equation by making the small angle approximation $\sin\theta \simeq \theta$. Some physical problems are better approximated by nonlinear oscillators such as

$$\dot{r} = 0, \qquad \dot{\theta} = -\omega + \alpha r + \cdots, \tag{7}$$

where $\alpha \neq 0$ is a constant. In such cases the solution curves will still be circles, since $\dot{r} = 0$, but the angular frequency, $-\omega + \alpha r + \cdots$, varies from circle to circle.

III. Pairs of Oscillators. Consider a pair of harmonic oscillators

$$\ddot{x} + \omega^2 x = 0, \qquad \ddot{y} + \mu^2 y = 0 \tag{8}$$

which as a system becomes

$$\begin{aligned} \dot{x} &= \omega u & \dot{y} &= \mu v \\ \dot{u} &= -\omega x & \dot{v} &= -\mu y. \end{aligned} \tag{9}$$

In polar coordinates it becomes

$$\begin{aligned} \dot{r} &= 0 & \dot{\rho} &= 0 \\ \dot{\theta} &= -\omega & \dot{\phi} &= \mathbf{x} - \mu \end{aligned} \tag{10}$$

and it admits integrals

$$I = x^2 + u^2 \qquad J = y^2 + v^2. \tag{11}$$

In many physical problems these equations are only the first approximation and the full system does not admit the two individual functions I and J as integrals but does admit the sum

$$E = I + J = x^2 + u^2 + y^2 + v^2 \tag{12}$$

or total energy as an integral. In this case the 3-sphere, $S^3 = E^{-1}(1)$, is an invariant set for the flow, (i.e. a solution which starts in this set stays in this set.) Think for example of a pea, of unit mass, rolling around in a bowl above the x-y plane with its minimum at the origin. If the only force acting on the pea is the constant force due to gravity, the potential, $V(x, y)$ of the pea would have a minimum at the origin. Let V have a Taylor expansion of the form $2V(x, y) = x^2 + y^2 + \cdots$. Then the equations of motion of the pea are $\ddot{x} = -V_x = -x + \cdots$, $\ddot{y} = -V_y = -y + \cdots$ and total energy is $(\dot{x}^2 + \dot{y}^2)/2 + (x^2 + y^2 + \cdots)/2$. Thus the linearized system at the minimum would be of the form (9) with $\omega = \mu = 1$.

Consider the general solution through $r_0, \rho_0, \theta_0, \phi_0$ at epoch $t = 0$. The solutions with $r_0 = 0$ and $\rho_0 > 0$ or $\rho_0 = 0$ and $r_0 > 0$ lie on circles and correspond to periodic solutions of period $2\pi/\mu$ and $2\pi/\omega$, respectively. These periodic solutions are special and are usually called the normal modes.

The set where $r = r_0 > 0$ and $\rho = \rho_0 > 0$ is an invariant torus for (9) or (10). Angular coordinates on this torus are θ and ϕ and the equations are

$$\dot{\theta} = -\omega \qquad \dot{\phi} = -\mu, \tag{13}$$

the standard linear equations on a torus.

If θ/μ is rational then $\omega = p\tau$ and $\mu = q\tau$ where p and q are relatively prime integers. In this case the solution of (13) through θ_0, ϕ_0 at epoch $t = 0$ is $\theta(t) = \theta_0 - \omega t$, $\phi(t) = \phi_0 - \mu t$ and so if $T = 2\pi/\tau$ then $\theta(T) = \theta_0 + p2\pi$ and $\phi(T) = \phi_0 + q2\pi$. That is, the solution is periodic with period T on the torus (see FIGURE 2a) and this corresponds to periodic solutions of (9).

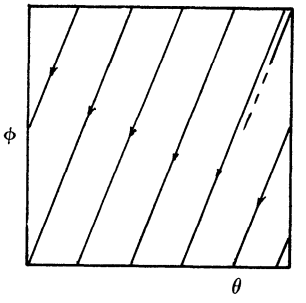


FIG. 2

If θ/μ is irrational then none of the solutions are periodic. In fact the solutions of (13) are dense lines on the torus (see FIGURE 2b) and this corresponds to the fact that the solutions of (9) are quasi-periodic but not periodic. See Coddington and Levinson [**3**] or Hale [**6**] for a discussion of the irrational flow on the torus.

We can use polar coordinates to introduce coordinates on the sphere provided we are careful to observe the conventions of polar coordinates: i) $r \geq 0$, ii) θ is defined modulo 2π and iii) $r = 0$ corresponds to a point. That is, if we start with the rectilinear strip $r \geq 0$, $0 \leq \theta \leq 2\pi$ (FIGURE 3a), identify the $\theta = 0$ and $\theta = 2\pi$ edges to get a half closed annulus (FIGURE 3b), and finally identify the circle $r = 0$ with a point, then we have a plane (FIGURE 3c).

Starting with the polar coordinates, r, θ, ρ, ϕ for R^4 we note that on the 3-sphere $E = r^2 + \rho^2 = 1$ so we can discard ρ and have $0 \leq r \leq 1$. We will use r, θ, ϕ as coordinates on S^3. Now, r, θ with $0 \leqslant r \leqslant 1$ are just polar coordinates for the closed unit disk. For each point of the open disk there is a circle with coordinate ϕ (defined mod 2π), but when $r = 1$, $\rho = 0$ so the circle collapses to a point over the boundary of the disk. The geometric model of S^3 is two solid cones with points on the boundary cones identified as shown in FIGURE 4. Through each point in the open unit disk with coordinates r, θ there is a line segment (the dashed line) perpendicular to the disk. The angular coordinate ϕ is measured on this segment; $\phi = 0$ is the disk, $\phi = \pi$ is the upper boundary cone, $\phi = -\pi$ is the lower boundary cone. Each point on the upper boundary cone with coordinates $r, \theta, \phi = \pi$ is identified with the point on the lower boundary cone with coordinate $r, \theta, \phi = -\pi$. From this model follows a series of interesting geometric facts.

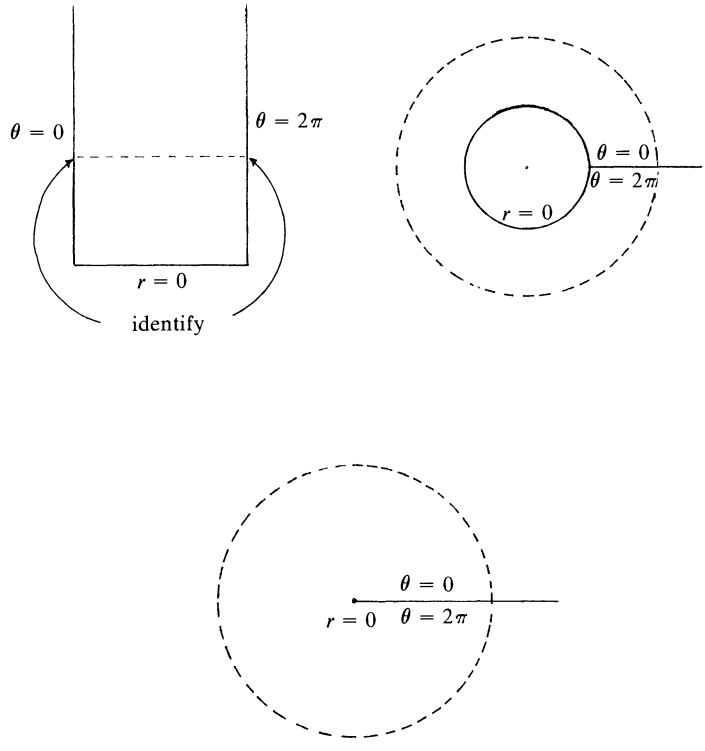


FIG. 3

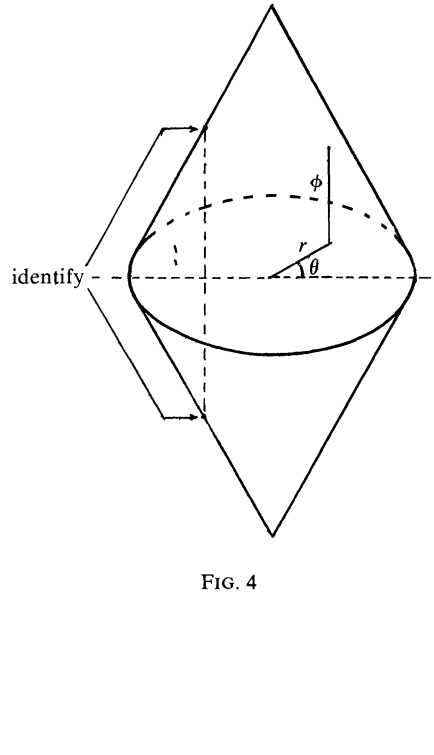


FIG. 4

For α, $0 < \alpha < 1$, the set where $r = \alpha$ is a 2-torus in the 3-sphere and for $\alpha = 0$ or 1 the set $r = \alpha$ is a circle. Since r is an integral for the pair of oscillators these tori and circles are invariant sets for the flow defined by the harmonic oscillators. The two circles, $r = 0, 1$ are periodic solutions, called the normal modes. The two circles are linked in S^3, i.e. one of the circles intersects a disk bounded by the other circle in an algebraically non-trivial way. The circle where $r = 1$ is the shaded disk in FIGURE 5 and the circle $r = 0$ intersects this disk once. It turns out that the number of intersections is independent of the bounding disk provided one counts the intersections algebraically. See Rolfsen [**10**].

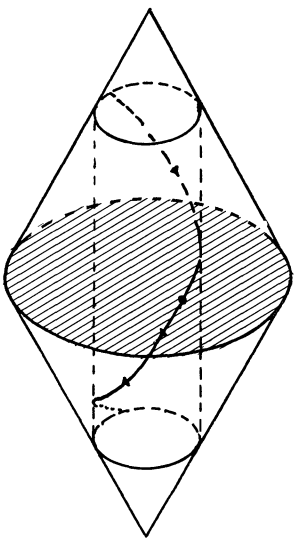


FIG. 5

Consider the special case when $\omega = \mu = 1$. In this case every solution is periodic and so its orbit is a circle in the 3-sphere. Other than the two special circles, on each orbit as θ increases by 2π so does ϕ. Thus each such orbit hits the open disk where $\phi = 0$ (the shade disk in FIGURE 5) in one point. We can identify each such orbit with the unique point where it intersects the disk. One special orbit meets the disk at the center and so we can identify it with the center. The other is the outer boundary circle which is a single orbit. When we identify this circle with a point, the closed disk with its outer circle identified with a point becomes a 2-sphere. Thus:

PROPOSITION. *The 3-sphere, S^3, is the union of circles. Any two of these circles are linked. The quotient space obtained by identifying each circle with a point, is a 2-sphere (the Hopf fibration of S^3).*

Let D be the open disk $\phi = 0$, the shaded disk in FIGURE 5. The union of all the orbits which meet D is a product of a circle and a 2-disk, so each point not on the special circle $r = 1$ lies in an open set which is the product of a 2-disk and a circle. By reversing r and ρ in the discussion given above the circle where $r = 1$ has a similar neighborhood. So locally the 3-sphere is the product of a disk and a

circle, but the sphere is not the product of a two manifold and a circle. (The sphere has trivial fundamental group but such a product would not. See Rolfsen [**10**].)

When $\omega = p$ and $\mu = q$, p and q relatively prime integers all solutions are periodic and the 3-sphere is again a union of circles but it is not locally a product near the special circles. Arnold [**1**] calls this decomposition into circles near the special circles the Seifert foliation and shows its connection with bifurcation theory. The nonspecial circles are p-q torus knots. They link p times with one special circle and q times with the other.

The linking statements follow by a slight extension of the ideas of the previous proposition. A p-q torus knot is a closed curve which wraps around the standard torus in R^3 in the longitudinal direction p times and in the meridional direction q times. If p and q are different from 1 the knot is nontrivial. We cannot go into details here, but FIGURE 6 shows that the 3-2 torus knot is the classical trefoil or clover-leaf knot. In FIGURE 6a the opposite sides of the squares are to be identified so that it represents a torus. The line of slope 2/3 in 6a wraps around the torus 3 times in one direction and 2 times in the other. This line is the important object so henceforth the sides of the torus will be suppressed in the drawings. Think of rolling the square in 6a into a cylinder by rolling the top half of 6a back of the bottom half. Now the line wraps over and under the cylinder and has two end points at each end of the cylinder. Join the end points with additional curves to get FIGURE 6b. Note that the crossing patterns for the curves in FIGURES 6b and 6c are the same and so represent the same curve. FIGURE 6c is the standard representation of the trefoil knot. See Rolfsen [**10**] for more information on knots.

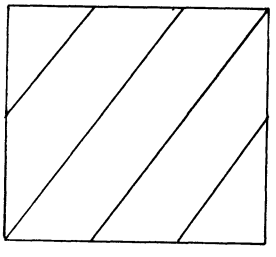


FIG. 6a

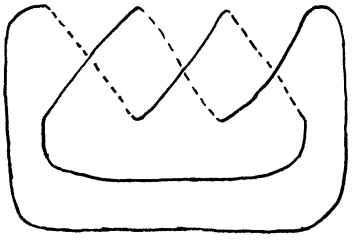


FIG. 6b

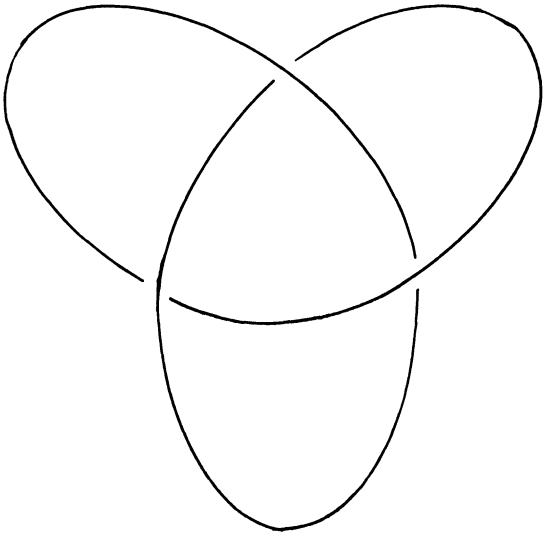


FIG. 6c

Sometimes one needs to consider nonlinear oscillators. A simple and somewhat typical nonlinear system is

$$\begin{aligned} \dot{r} &= 0 & \dot{\rho} &= 0 \\ \dot{\theta} &= \omega + \alpha r & \dot{\phi} &= \mu + \beta\rho, \end{aligned} \tag{14}$$

where α and β are constants. In this case we still have the special circles, the decomposition into tori, r = constant as described above, and the flow on the tori is still linear. However, now the slope of the linear flow on the tori varies from torus to torus. Sometimes all orbits are dense and sometimes all orbits are periodic.

IV. The Kepler Problem. As stated above, pairs of linear oscillators are often the first approximation to interesting nonlinear physical problems, but also they completely solve the Kepler problem of celestial mechanics. Consider a body of mass M fixed at the origin in the plane and another free body of mass m whose position in the plane is given by the vector q. Assume that the free body is attracted to the fixed body by the force given by Newton's law of gravity. Then the equation of motion for the free body is

$$\ddot{q} = \frac{-GMq}{\|q\|^3}, \tag{15}$$

where G is the universal gravitational constant. By changing the time scale if necessary the constant GM can be taken as 1. There are many ways to solve this system of differential equations, but I think one of the most informative ways is the regularization procedure of Levi-Civita. This method removes the singularity at the origin and reduces the equation (15) to a pair of harmonic oscillators with equal frequencies.

Consider q as a complex number, so multiplication makes sense and the norm $\|\cdot\|$ can be replaced with absolute value $|\cdot|$. Since the bounded orbits of the problem occur for negative energy, write the energy integral as

$$|\dot{q}|^2 - 2/|q| = -C, \tag{16}$$

so C is the negative of the total energy. The Levi-Civita change of coordinates is $q = w^2$ together with a change of time $dt = 4|w|^2\, d\tau$. The transformation from w to q is a 2 to 1 transformation except at the origin. After some careful computations the equation of motion in the new coordinates becomes

$$\frac{d^2w}{d\tau^2} + 4Cw = 0, \tag{17}$$

a pair of harmonic oscillators with frequencies $4C$ when $C > 0$. The computations are carried out in full in Szebehely [9].

Thus the flow of the Kepler problem (15) on a negative energy level is the 2 to 1 projection of the flow on the 3-sphere defined by (17) where the projection is the squaring map. In the new coordinates $w = 0$ corresponds to a collision of the two bodies but the equations (17) do not have a singularity there. Thus these coordinates 'regularize the collision singularity'.

REFERENCES

1. V. I. Arnold, Geometrical Methods in the Theory of Ordinary Differential Equations, Springer-Verlag, New York, 1983.
2. G. D. Birkhoff, The restricted problem of three bodies, *Rend. Circ. Mat. Palermo*, 39 (1915) 265–334.
3. E. Coddington and N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.
4. C. C. Conley, On some new long periodic solutions of the plane restricted three body problem, *Comm. Pure and Appl. Math.*, XVI (1963) 449–467.
5. E. R. Fadell and P. H. Rabinowitz, Generalized cohomological index theories for Lie group actions with applications to bifurcation questions for Hamiltonian systems, *Inv. Math.*, 45 (1978) 134–174.
6. J. K. Hale, Ordinary Differential Equations, Wiley, New York, 1969.
7. H. Ito, Long periodic orbits near an equilibrium and the averaging method in higher order resonances, *J. Dif. Eqs.*, 51 (1984) 232–253.
8. J. Moser, Periodic orbits near an equilibrium and a theorem of Weinstein, *Comm. Pure and Appl. Math.*, XXIX (1976) 727–747.
9. V. Szebehely, Theory of Orbits, Acad. Press, New York, 1967.
10. D. Rolfsen, Knots and Links, Publish or Perish, Berkeley, CA, 1976.
11. A. Weinstein, Normal modes for non-linear Hamiltonian systems, *Inv. Math.*, 20 (1973) 47–57.

Game of Cards, Dynamical Systems, and a Characterization of the Floor and Ceiling Functions

P. EISELE and K. P. HADELER



PETER EISELE obtained his diploma in Mathematics from the University of Tübingen in 1987 and is now working for a doctoral degree in mathematics. His field is Structural Probability Theory and Measure Theory.



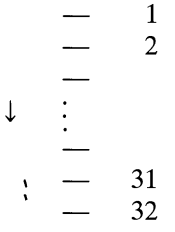
K. P. HADELER obtained his doctoral degree (1965) and his Habilitation (1967) in Mathematics from the University of Hamburg. He has been Vis. Assoc. Professor at the University at Minnesota, 1969, and Assoc. Professor at the University of Erlangen 1970–71. Since 1971 he is Professor for Mathematics and Biomathematics at the University of Tübingen. His fields are Differential Equations, Applied Mathematics, Mathematical Biology.

1. Card games and dynamical systems. In the past thirty years the abstract qualitative theory of ordinary differential equations and of difference equations has been cast into the theory of dynamical systems with continuous time and with discrete time, respectively. In general, in the theory of dynamical systems the state space is a smooth finite-dimensional manifold. Depending on the dimension of this manifold, its topological structure, and whether time is continuous or discrete, the qualitative behavior of trajectories may be more or less complicated. For continuous time, in dimension one, every bounded trajectory converges to a stationary point. On the sphere of dimension two, the theorem of Poincaré and Bendixson ensures, that (for vector fields with finitely many hyperbolic stationary points) limit sets are points or periodic orbits or cyclic chains of heteroclinic orbits connecting saddle points [7]. For dimensions greater than three the possible behaviors of trajectories are very complicated. Dynamical systems with discrete time already show very complicated behavior in one dimension (comparable to differential equations in three-dimensional space). However, for monotone systems every bounded trajectory converges to a stationary point. In nonmonotone systems cycles appear, typically of order 2 or a power of 2 (period-doubling, Feigenbaum sequence [1], [2]) or more complicated behavior (Sharkovskij families of periodic orbits, "chaotic" behavior [2], [6]). In the following we shall study even more discrete dynamical systems, namely systems on the set of integers $\mathbb{Z}$. Of course, supplied only with the discrete topology, one loses much structure, but general questions about trajectories and their asymptotic behavior, about the structure of limit sets, the number of periodic orbits etc. remain the same. One is even

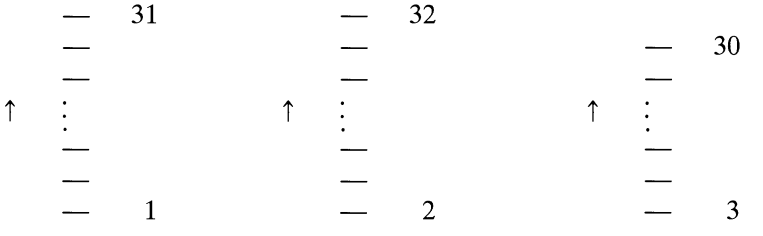
rewarded by the appearance of the new concept of stop time, which in some sense replaces the rate of convergence to a stationary point. The most famous dynamical system on $\mathbb{Z}$ is the $3x + 1$-game, a very simple recursion formula for which the convergence problem is unsolved for more than fifty years despite many efforts (see [5] and the references therein). Here we study a much simpler class of dynamical systems on $\mathbb{Z}$ which are related to the floor and ceiling functions. The point of depart towards these systems is located in a quite different field of mathematics, it is a combinatorial problem related to a certain game of cards. This game seems, indeed, to be mathematical folklore, and is described in a book of that title by Gančev et al. [3], for a special case where certain divisibility conditions are satisfied (which are not satisfied for the practical cases of 32 or 52 cards). Nevertheless in [3] a recurrence relation for a bound on the stop time in terms of the floor function is given. We observe that this recurrence relation can be solved by using an obvious functional equation for the floor function. Even more, using the functional equation, the exact stop time can be determined in the general case. What seems equally interesting, is that the functional equation characterizes the floor and ceiling functions. Of course the result on card games and on recurrence relations could be formulated without reference to dynamical systems, but the language of dynamical systems seems to be most appropriate since the discrete analogon of bifurcation to periodic orbits occurs.

Consider a well-known playing card trick. Player A has a deck of 32 cards. Player B remembers a card, A deals the cards, one by one, cyclically into three packs of 11, 11, 10 cards. B indicates which pack contains his card. A collects the cards into one deck, whereby he puts the indicated pack between the other two in such a way that a pack with 11 cards goes on top, deals again, and so on. After four steps of this procedure, the remembered card is known to A, it assumes the seventeenth position in the deck. There are several variants of this trick depending on how the cards are collected into one deck and how they are redistributed. We shall consider generalizations of two versions of this game which essentially differ in their mathematical properties.

Now we give an exact description of the first of these procedures. The cards are in the deck, face down, numbered from 1 to 32



They are distributed into three packs, face up,



If the indicated pack is the first, then the second is taken first, reversed, then the first is packed underneath, reversed, finally the third, to give

$$\begin{array}{cc} & — \quad 2 \\ \downarrow & \vdots \\ & — \quad 32 \\ & — \quad 1 \\ \downarrow & \vdots \\ & — \quad 31 \\ & — \quad 3 \\ \downarrow & \vdots \\ & — \quad 30 \end{array}$$

a face down deck with the 25th card in position 20. It is easy to see that for each card of the indicated pack the new position x_1 is obtained from the previous position x_0 by the formula

$$x_1 = 11 + \left\lceil \frac{x_0}{3} \right\rceil = \left\lceil \frac{32}{3} \right\rceil + \left\lceil \frac{x_0}{3} \right\rceil, \tag{1}$$

where $\lceil x \rceil$, for real x, denotes the ceiling function, i.e., $\lceil x \rceil$ is the smallest integer number not less than x. (The floor, or ENTIER function, i.e., the largest integer not greater than x, will be denoted by $\lfloor x \rfloor$, observe $\lceil x \rceil = -\lfloor -x \rfloor$.) Hence the card game induces a dynamical system on the set $\{1, 2, \ldots, 32\}$. The system depends on a parameter, the number of packs, in this case 3. In a general version considered here the game depends on three parameters N, b, p. There are N cards in the deck, these are numbered from 1 to N from top, face down. They are distributed into b packs, face up. After reversing the packs, p cards are put on top of the indicated pack and the remaining cards are put underneath. Then instead of (1) the new position x_1 of the card x_0 is given by

$$x_1 = p + \left\lceil \frac{x_0}{b} \right\rceil. \tag{2}$$

Hence the parameter N is not essential. The general procedure can be interpreted as a dynamical system on $\mathbb{Z}$

$$x_{n+1} = f(x_n), \qquad n = 0, 1, 2, \ldots, \tag{3}$$

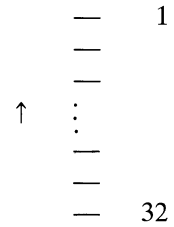
where $f\colon \mathbb{Z} \to \mathbb{Z}$ is given by

$$f(x) = a + \left\lceil \frac{x}{b} \right\rceil, \tag{4}$$

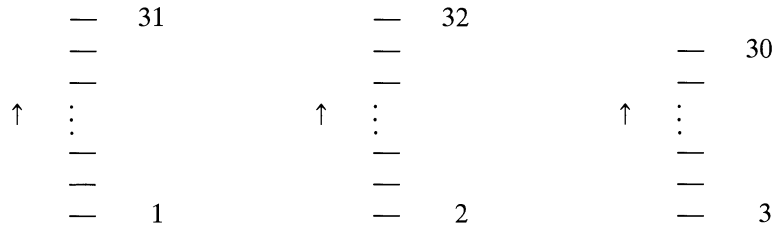
for given $a \in \mathbb{Z}$, $b \in \mathbb{N} \setminus \{1\}$.

In the second variant the player will deal cards from the top of a face up deck and will never reverse them. It turns out that this apparently simpler procedure is

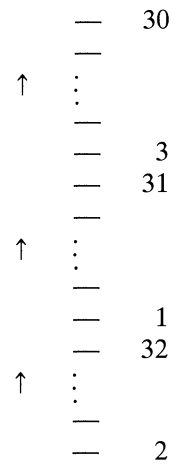
mathematically more difficult to analyze. The arrangement



becomes



Now player A puts a pack of 11 underneath, the indicated pack in the middle, and the last one on top. If the indicated pack is the first, then the result is



The 25th card moves into position 13. The distribution procedure reverses the order in each pack, and this reversed order is retained when the packs are collected into the deck. We shall see that this reversal leads to a different behavior of the iteration. The general procedure is as follows. A deck of N cards, face up, is numbered from 1 to N. The cards are distributed from the top into b packs. The pack that contains the desired card is taken first, then p cards are put underneath and the remaining cards on top. Then the new position of the card at x_0 is given by

$$x_1 = N + 1 - p - \left\lceil \frac{x_0}{b} \right\rceil. \tag{5}$$

In the above example $N = 32$, $b = 3$, $p = 11$. Hence, games of cards with reversal lead to dynamical systems of the form

$$x_{n+1} = f(x_n), \qquad n = 0, 1, 2, \ldots, \tag{3}$$

$$f(x) = a - \left\lceil \frac{x}{b} \right\rceil. \tag{6}$$

Some other variants of the card games may at first sight lead to recursion formulas containing the floor instead of the ceiling function. However, notice that the dynamical systems

$$x_{n+1} = a \pm \left\lceil \frac{x_n}{b} \right\rceil \tag{7}$$

and

$$y_{n+1} = -a \pm \left\lfloor \frac{y_n}{b} \right\rfloor \tag{8}$$

are equivalent (conjugated): put $x_n = -y_n$ and use $\lceil x \rceil = -\lfloor -x \rfloor$. Thus one can always replace the floor by the ceiling function and vice versa, but one cannot change the minus sign into a plus sign. We shall see later that the sign distinguishes two very different classes of problems.

The following questions arise: Are there stationary points of the dynamical systems (3)(4) and (3)(6) (fixed points of the mapping f)? How many? Does every trajectory converge to a stationary point? To which stationary point does a given trajectory converge? Are there cycles? What is the stop time for given x_0?

We shall prove that in the dynamical system (3)(4) there are at most two stationary points, every trajectory converges to one of these. We give an explicit expression for the stop time. As a consequence we know for the general game of cards how many steps are necessary to determine the unknown card. For the system (3)(6) we can distinguish the cases with stationary points from those with periodic orbits of order two. The key fact of the trick with 32 cards and 3 packs is that, in the first version, every trajectory with $x_0 \in \{1, \ldots, 32\}$ reaches the only stationary point $\bar{x} = 17$ after at most 4 steps.

2. Dynamical systems on $\mathbb{Z}$. First we consider a general dynamical system (3).

LEMMA 1. *Let $f\colon \mathbb{Z} \to \mathbb{Z}$ be any function with the following properties.*

i) $$x \leqslant y \Rightarrow f(x) \leqslant f(y).$$

ii) $$\exists \bar{x}, \quad \underline{x} \in \mathbb{Z} \quad \text{such that } f(x) \leqslant x \text{ holds for } x \geqslant \bar{x}, \quad f(x) \geqslant x \text{ holds for } x \leqslant \underline{x}.$$

Then every trajectory of the dynamical system (3) *is bounded and converges to a stationary point.*

Proof. If x_0 is such that $x_1 \geqslant x_0$ then $\{x_n\}$ is nondecreasing. If x_0 is such that $x_1 \leqslant x_0$ then $\{x_n\}$ is nonincreasing. Hence $\{x_n\}$ is nondecreasing for $x_0 \ll 0$ and nonincreasing for x_0 large. Hence for all x_0 the sequence $\{x_n\}$ is monotone and bounded. Since it is integer valued, it becomes eventually constant, and the limit is trivially a stationary point. □

For typographical convenience we write

$$\vartheta = \lceil\ \rceil, \qquad \overline{\vartheta} = \lfloor\ \rfloor. \tag{9}$$

The function $\vartheta\colon \mathbb{R} \to \mathbb{Z}$, (and also $\overline{\vartheta}\colon \mathbb{R} \to \mathbb{Z}$) has the following properties (see Sect. 3)

i) *For all* $x \in \mathbb{R}$, $c \in \mathbb{Z}$ *holds*

$$\vartheta(x + c) = \vartheta(x) + c. \tag{10}$$

ii) *For all* $x \in \mathbb{R}$, $n \in \mathbb{N}$ *holds*

$$\vartheta\left(\frac{1}{n}\vartheta(nx)\right) = \vartheta(x). \tag{11}$$

iii) ϑ *is nondecreasing.* (12)

Now we start the investigation of the dynamical system

$$x_{n+1} = f(x_n), \qquad n = 0, 1, 2, \ldots, \tag{3}$$

$$f(x) = a + \vartheta(x/b). \tag{13}$$

THEOREM 2. *Consider the dynamical system* (3)(13).

i) *If* $a \not\equiv 0 \bmod(b - 1)$ *then*

$$\bar{x} = \vartheta\left(\frac{ab}{b - 1}\right) \tag{14}$$

is the only stationary point, and every trajectory converges monotonely to $\bar{x}$.

ii) *If* $a \equiv 0 \bmod(b - 1)$ *then*

$$\bar{x}_1 = \frac{ab}{b - 1}, \qquad \bar{x}_2 = \bar{x}_1 + 1 \tag{15}$$

are the only stationary points, and every trajectory converges monotonely to the nearest stationary point.

Proof. Let

$$a = \mu(b - 1) + \rho, \qquad 0 \leqslant \rho < b - 1$$

and assume that

$$x = mb + r, \qquad 0 \leqslant r < b, \qquad m \in \mathbb{Z},$$

is a stationary point. Then

$$mb + r = \mu(b - 1) + \rho + m + \vartheta(r/b)$$

or

$$(m - 1)(b - 1) = \rho - r + \vartheta(r/b).$$

Since $\rho < b - 1$ and $r < b$, this inequality implies $m = \mu$ and

$$r - \vartheta(r/b) = \rho.$$

Case 1: $\rho > 0$, i.e., $a \equiv 0 \bmod(b - 1)$. Then $r \neq 0, 1$ and $r = \rho + 1$, and we obtain

$$x = \mu b + \rho + 1 = a + \mu + 1 = a + \frac{a}{b - 1} + \frac{b - 1 - \rho}{b - 1}$$

$$= a + \vartheta\left(\frac{a}{b - 1}\right) = \vartheta\left(\frac{ab}{b - 1}\right) = \bar{x}.$$

Case 2: $\rho = 0$, i.e., $a \equiv 0 \bmod(b-1)$. Then $r = 0$ or $r = 1$. With $r = 0$ we obtain

$$x = \mu b = a + \mu = a + \frac{a}{b-1} = \frac{ab}{b-1} = \bar{x}_1.$$

With $r = 1$ we obtain

$$x = \mu b + 1 = \bar{x}_1 + 1 = \bar{x}_2.$$

One checks immediately that the obtained numbers are stationary points. The convergence is evident from Lemma 1. □

Since f as defined in (13) has the monotonicity properties required in Lemma 1 the convergence problem becomes so simple, quite in contrast, say, to the $3x + 1$ map (cf. [5]). Also in the theory of monotone dynamical systems with continuous time the limit sets have a particularly simple structure (cf. [4]).

THEOREM 3. *Consider f as in* (13). *Then for $x \in \mathbb{Z}$*

$$f^{(k)}(x) = \vartheta\left(\frac{ab}{b-1} + \left(x - \frac{ab}{b-1}\right)b^{-k}\right), \qquad k = 0, 1, 2, \ldots . \tag{16}$$

Proof. For $k = 0$ the assertion is trivial. Assume (15) is true for some k. Then

$$\begin{aligned} f^{(k+1)}(x) &= a + \vartheta\left(\frac{f^{(k)}(x)}{b}\right) \\ &= a + \vartheta\left(\frac{1}{b}\vartheta\left(\frac{ab}{b-1} + \left(x - \frac{ab}{b-1}\right)b^{-k}\right)\right) \\ &= a + \vartheta\left(\frac{a}{b-1} + \left(x - \frac{ab}{b-1}\right)b^{-k-1}\right) \\ &= \vartheta\left(a + \frac{a}{b-1} + \left(x - \frac{ab}{b-1}\right)b^{-(k+1)}\right) \\ &= \vartheta\left(\frac{ab}{b-1} + \left(x - \frac{ab}{b-1}\right)b^{-(k+1)}\right). \end{aligned}$$

□

Remark. The formula (16) implies convergence to a stationary solution for $k \to \infty$. Furthermore the formula exhibits, for each $x \in \mathbb{Z}$, the limit of the trajectory, if the discontinuities of ϑ are carefully inspected.

For given x_0 the stop time is the smallest n such that x_n is a point of a periodic orbit. For large $|x|$ the difference between $a + \lceil x/b \rceil$ and x/b is relatively small (as compared to $|x|$). Therefore the stop time is close to $(\log|x|)/\log b$ for large $|x|$. In fact the stop time can be given explicitly.

COROLLARY 4. *Consider the dynamical system* (3)(13). *Let a, b be fixed. Let $\tau = \tau(x)$ be the stop time for the trajectory starting at x. Let $B = ab/(b-1)$. Then*

i) *if $b - 1$ does not divide a and $x > \vartheta(ab/(b-1))$*

$$\tau(x) = \vartheta\left(\log\frac{x - B}{\vartheta(B) - B}/\log b\right)$$

ii) *if* $b-1$ *does not divide* a *and* $x < \vartheta(ab/(b-1))$

$$\tau(x) = \bar{\vartheta}\left(\log \frac{B-x}{1-(\vartheta(B)-B)} / \log b\right) + 1.$$

iii) *if* $b-1$ *divides* a *and* $x > ab/(b-1) + 1$

$$\tau(x) = \vartheta\left(\frac{\log(x-B)}{\log b}\right).$$

iv) *if* $b-1$ *divides* a *and* $x < ab/(b-1)$

$$\tau(x) = \bar{\vartheta}\left(\frac{\log(B-x)}{\log b}\right) + 1.$$

The proof follows from Theorem 3 by solving for k. □

Next we consider the dynamical system

$$x_{n+1} = f(x_n), \qquad n = 0, 1, 2, \ldots, \tag{3}$$

$$f(x) = a - \vartheta\left(\frac{x}{b}\right). \tag{17}$$

Now f is not monotonely increasing and the system is somewhat more complicated. Let $F = f \circ f$. Then from $f(x) = -(\vartheta(x/b) - a) = -\vartheta(x/b - a)$

$$F(x) = (f \circ f)(x) = a - \vartheta\left(\frac{-1}{b}\vartheta\left(\frac{x}{b} - a\right)\right)$$

$$= a + \bar{\vartheta}\left(\frac{1}{b}\vartheta\left(\frac{x}{b} - a\right)\right). \tag{18}$$

THEOREM 5. *Every trajectory of the system* (3)(17) *becomes eventually constant or evolves into a cycle of order* 2. *Furthermore*

i) *If* $a \not\equiv 1 \bmod(b+1)$, *then the function* f *has exactly one fixed point*

$$\bar{x} = \bar{\vartheta}\left(\frac{ab}{b+1}\right). \tag{19}$$

ii) *If* $a \equiv 1 \bmod(b+1)$, *then* $f \circ f$ *has exactly two fixed points, namely,*

$$\bar{x}_1 = \frac{b(a-1)}{b+1}, \qquad \bar{x}_2 = \bar{x}_1 + 1 \tag{20}$$

which form an attracting two-cycle of the dynamical system, and f *has no fixed point.*

Proof of Theorem 5. The function F is monotonely increasing and satisfies, since $b \geq 2$, the hypothesis of Lemma 1. Hence, for each trajectory $\{x_n\}$ the subsequences $\{x_{2k}\}$ and $\{x_{2k+1}\}$ converge, and the limits are trivially fixed points of F.

1) Let

$$a = \mu(b+1) + \rho, \qquad 0 \leq \rho < b+1.$$

Assume that

$$x = mb + r, \qquad 0 \leq r < b$$

is a stationary point. Then

$$mb + r = \mu(b + 1) + \rho - m - \vartheta(r/b)$$

or

$$(m - \mu)(b + 1) = \rho - r - \vartheta(r/b).$$

This equality implies $m = \mu$ and

$$r + \vartheta(r/b) = \rho.$$

Hence $\rho \neq 1$, i.e., $a \not\equiv 1 \bmod(b + 1)$.

Case 1: $\rho = 0$. Then $r = 0$ and

$$x = \mu b = a - \mu = a - \frac{a}{b + 1} = \frac{ab}{b + 1} = \bar{x}.$$

Case 2: $\rho > 1$, then $r + 1 = \rho$, and

$$x = \mu b + \rho - 1 = a - \mu - 1 = a - \left(\frac{a}{b + 1} - \frac{\rho}{b + 1} + 1\right)$$

$$= a - \vartheta\left(\frac{a}{b + 1}\right) = \bar{\vartheta}\left(\frac{ab}{b + 1}\right) = \bar{x}.$$

On the other hand, if $\rho \neq 1$, then $\bar{x}$ is a stationary point. Hence i) is proved.

2) It remains to analyze the case where there is no stationary point or, equivalently, $\rho = 1$. A fixed point of F corresponds to a pair x_1, x_2 such that

$$\begin{aligned} x_1 &= a - \vartheta(x_2/b), \\ x_2 &= a - \vartheta(x_1/b). \end{aligned} \tag{21}$$

Assume

$$x_1 = mb + r, \qquad x_2 = nb + s, \qquad 0 \leqslant r, s < b.$$

Case 1: $r, s > 0$. From

$$\begin{aligned} mb + r &= a - n - 1 \\ nb + s &= a - m - 1 \end{aligned}$$

it follows that $(m - n)(b - 1) = s - r$. Thus $b - 1$ divides $s - r$. But $|s - r| < b - 1$. Hence $s - r = 0$ and $m = n$. The point $x_1 = x_2$ is a fixed point of f. Contradiction.

Case 2: $r = s = 0$. From

$$\begin{aligned} mb &= a - n \\ nb &= a - m \end{aligned}$$

it follows that $(m - n)(b - 1) = 0$, thus $m = n$, and $x_1 = x_2$ is a fixed point of f. Contradiction.

Case 3: $r = 0$, $s > 0$ (or vice versa). From

$$\begin{aligned} mb &= a - n - 1, \\ nb + s &= a - m \end{aligned}$$

it follows that $(m - n)(b - 1) = s - 1$. Hence $b - 1$ divides $s - 1$. But $s - 1 \leqslant b - 2$. Therefore $s = 1$ and $m = n$, and $a = m(b + 1) + 1$. The numbers $x_1 = mb$ and $x_2 = mb + 1$ satisfy equations (21) and are equal to $\bar{x}_1$ and $\bar{x}_2$. □

Unfortunately, we have not been able to find an "explicit" solution formula or to determine the stop time.

Remark. Since the systems (7) and (8) are equivalent, all results can be carried over to the case where $\vartheta = \lfloor\ \rfloor$ (and, of course, $\overline{\vartheta} = \lceil\ \rceil$). In particular, in Theorem 2,ii, one has $\bar{x}_2 = ab/(b-1)$ and $\bar{x}_1 = \bar{x}_2 - 1$, Theorem 3 remains as it is, in Corollary 4 one obtains only slightly different expressions. Finally, in Theorem 5 the special case to be observed is $a \equiv b \bmod(b+1)$, where $\bar{x}_2 = b(a+1)/(b+1)$ and $\bar{x}_1 = \bar{x}_2 - 1$.

To prove Theorem 3 one only needs the properties (9), (10). Here some hope may arise that a more general class of functions could be considered. See, however, the next section.

3. Characterization of the floor and ceiling functions.

LEMMA 6. *The function* $\vartheta(x) = \lceil x \rceil$ *has the properties* (9)(10)(11).

Proof. (9)(11) are evident. (10) is equivalent to

$$\vartheta(\vartheta(x)/n) = \vartheta(x/n). \tag{22}$$

Assume $x = mn + r$, $0 \leqslant r < n$, $m \in \mathbb{Z}$. For $r = 0$ the assertion is evident. Now let $r > 0$. Then

$$\begin{aligned}\vartheta(\vartheta(x)/n) &= \vartheta(\vartheta(mn + r)/n)\\ &= \vartheta((mn + \vartheta(r))/n)\\ &= \vartheta(m + \vartheta(r)/n) = m + 1.\end{aligned}$$

On the other hand

$$\vartheta(x/n) = \vartheta(m + r/n) = m + 1. \qquad \square$$

LEMMA 7. *If a function* ϑ *satisfies* (9)(10)(11) *then the function* $\overline{\vartheta}$, *as defined by* $\overline{\vartheta}(x) = -\vartheta(-x)$, *satisfies* (9)(10)(11).

Proof.

$$\overline{\vartheta}(x + c) = -\vartheta(-x - c) = -\vartheta(-x) + c = \overline{\vartheta}(x) + c,$$

$$\begin{aligned}\overline{\vartheta}\left(\frac{1}{n}\overline{\vartheta}(nx)\right) &= -\vartheta\left(-\frac{1}{n}\overline{\vartheta}(nx)\right) = -\vartheta\left(\frac{1}{n}\vartheta(-nx)\right)\\ &= -\vartheta(-x) = \overline{\vartheta}(x).\end{aligned}$$

$\square$

Finally we show that the conditions (9)(10) do not lead to any generality. The proof is surprisingly complicated (unless a simple argument has escaped our attention) and for this reason it will be given here.

THEOREM 8. *Assume* $\vartheta: \not\mathbb{Q} \to \mathbb{Z}$ *has the following properties*

i) *For all* $x \in \not\mathbb{Q}$, $c \in \mathbb{Z}$ *holds*

$$\vartheta(x + c) = \vartheta(x) + c.$$

ii) *For all* $x \in \not\mathbb{Q}$, $n \in \mathbb{N}$ *holds*

$$\vartheta\left(\frac{1}{n}\vartheta(nx)\right) = \vartheta(x)$$

or, equivalently,

$$\vartheta\left(\frac{1}{n}\vartheta(x)\right) = \vartheta\left(\frac{1}{n}x\right).$$

Then either $\vartheta(x) = \lceil x \rceil$ *for all* $x \in \not\emptyset$ *or* $\vartheta(x) = \lfloor x \rfloor$ *for all* $x \in \not\emptyset$.

Proof.

1. For all $x \in \not\emptyset$, $\vartheta(\vartheta(x)) = \vartheta(x)$, i.e., all images of ϑ are fixed points.
2. Assume $\vartheta(0) = z \in \mathbb{Z}$. Then

$$\vartheta(0) = z = \vartheta(z) = \vartheta(0 + z) = \vartheta(0) + z,$$

hence $z = 0$,

$$\vartheta(0) = 0. \tag{23}$$

Thus for $c \in \mathbb{Z}$

$$\vartheta(c) = c. \tag{24}$$

3. Assume $\vartheta(1/2) = z \in \mathbb{Z}$. For $k \in \mathbb{N}_0$

$$z = \vartheta\left(\frac{1}{2k+1}\vartheta\left(\frac{2k+1}{2}\right)\right) = \vartheta\left(\frac{1}{2k+1}\vartheta\left(k + \frac{1}{2}\right)\right)$$
$$= \vartheta\left(\frac{k + \vartheta\left(\frac{1}{2}\right)}{2k+1}\right) = \vartheta\left(\frac{k+z}{2k+1}\right).$$

If $z < 0$ then put $k = -z$ to obtain $z = \vartheta(0) = 0$. If $z > 1$ then put $k = z + 1$ to obtain $z = \vartheta(1) = 1$. Hence

$$\vartheta\left(\frac{1}{2}\right) = 0 \text{ or } \vartheta\left(\frac{1}{2}\right) = 1. \tag{25}$$

4. Assume $\vartheta(1/2) = 0$. Claim:

$$\vartheta\left(\frac{1}{n}\right) = 0 \text{ for } n \in \mathbb{N}, \qquad n > 1. \tag{26}$$

Proof by induction. The assertion is true for $n = 2$. Assume it is true for some $n - 1 \geqslant 2$. Then

$$1 = \vartheta\left(1 + \frac{1}{n-1}\right) = \vartheta\left(\frac{n}{n-1}\right)$$

and thus

$$\vartheta\left(\frac{1}{n}\right) = \vartheta\left(\frac{1}{n}\vartheta\left(\frac{n}{n-1}\right)\right) = \vartheta\left(\frac{1}{n-1}\right) = 0.$$

5. Still assume $\vartheta(1/2) = 0$. Claim:

$$\vartheta\left(\frac{k}{n}\right) = 0 \text{ for } n, k \in \mathbb{N}, \qquad k < n. \tag{27}$$

Proof by induction over k. The assertion is true for $k = 1$. Assume

$$\vartheta\left(\frac{\nu}{n}\right) = 0 \text{ for } \quad \nu = 1, \ldots, k-1 \text{ and } n > \nu.$$

Let $n > k$ and $n = (m-1)k + r$, $0 \leqslant r < k$, thus $m > 1$. If $r = 0$ then k divides n and $\vartheta(k/n) = \vartheta(1/m - 1) = 0$. If $0 < r < k$, then

$$m = \frac{n + k - r}{k},$$

$$\frac{mk}{n} = 1 + \frac{k - r}{n},$$

and thus, by assumption

$$\vartheta\left(\frac{mk}{n}\right) = 1 + \vartheta\left(\frac{k - r}{n}\right) = 1.$$

Now

$$\vartheta\left(\frac{k}{n}\right) = \vartheta\left(\frac{1}{m}\vartheta\left(\frac{mk}{n}\right)\right) = \vartheta\left(\frac{1}{m}\right) = 0.$$

Hence it is proved that $\vartheta(1/2) = 0$ implies $\vartheta(x) = \lfloor x \rfloor$ for all $x \in \not\emptyset$.

6. Now assume that $\vartheta(1/2) = 1$. Then $\overline{\vartheta}(1/2) = 0$, thus $\overline{\vartheta}(x) = \lfloor x \rfloor$ for all $x \in \not\emptyset$, $\vartheta(x) = \lceil x \rceil$ for all $x \in \not\emptyset$. □

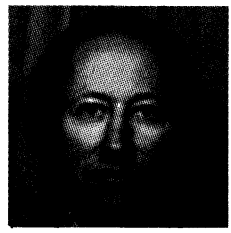
The authors thank a referee for useful suggestions.

REFERENCES

1. P. Collet, and J.-P. Eckmann, Iterated Maps on the Interval as Dynamical Systems, Birkhäuser Verlag, 1980.
2. R. L. Devaney, Chaotic Dynamical Systems, Benjamin Publ. 1987.
3. I. Gančev, K. Čimev, and I. Stoyanov, Matematiceski folklor, Narodna Prosveta, Sofia, 1987.
4. M. Hirsch, Systems of ordinary differential equations which are cooperative or competitive I. Limit sets, *SIAM J. Math. Anal.*, 13, (1982) 167–179.
5. J. C. Lagarias, The $3x + 1$ problem and its generalizations, *Amer. Math. Monthly*, 92 (1985) 3–23.
6. T. Li, J. A. Yorke, Period three implies chaos. *Amer. Math. Monthly*, 82 (1975) 985–92.
7. J. Palis, and W. di Melo, Geometric Theory of Dynamical Systems, Springer-Verlag, 1982.

Decomposition Theory in the Teaching of Elementary Linear Algebra

R. R. London and H. P. Rogosinski, *University College of Swansea, Singleton Park, Swansea SA2 8PP, U.K.*



Richard London received his Ph.D. in mathematics from Imperial College in the University of London in 1968. After a one year fellowship at the University of Kent he became a lecturer at the University College of Swansea in Wales. He has published papers in analysis, probability theory and linear algebra.



Peter Rogosinski received his Ph.D. in mathematics from the University of Cambridge in 1958. He was a lecturer at the University of Birmingham from 1958 to 1964, and since 1964 has been on the staff in the Mathematics Department at the University College of Swansea in Wales. His research interests are mainly in classical and functional analysis.

The standard curriculum for a first course in linear algebra will normally contain a proof of the Cayley-Hamilton theorem, a discussion of criteria for the diagonalizability of complex square matrices and perhaps some mention of the functional calculus for such matrices. In this article we describe a decomposition theory from which these topics are easily developed, and which we think is sufficiently simple for inclusion at this level. The theory and its applications show how much can be achieved that is based purely on the use of elementary polynomial algebra. In particular we have been able to omit Jordan canonical forms, and yet retain areas to which this topic applies.

The content of the theory is not new (see, e.g. [**4**, pp. 220–223]), but it seems to us that it has never been presented at an elementary level. Our approach *is* elementary, and we find that our students have accepted it readily. We therefore propose it now to teachers of linear algebra as a possible curriculum development. We also present a possibly novel approach to the functional calculus for square matrices, based on decomposition theory, which we have found successful in our teaching.

The decomposition theory is outlined in §1, drawing from [**4**]. In the following sections, we explain how the topics mentioned above are developed by means of this theory. The definition of $f(A)$ in §3 is suggested by Taylor's theorem. Indeed, by decomposition theory, A can be expressed as the sum of a diagonalizable matrix D with a nilpotent matrix N, and the expansion of $f(D + N)$ about D is a *finite* sum due to the nilpotency of N. This definition, although equivalent to that given by other authors has, to the best of our knowledge, not been previously presented in the literature. We conclude in the final section with a discussion of some interesting examples used to illustrate the theory and its applications.

We should like to thank the referee for some very helpful comments concerning the presentation of the paper.

1. Throughout, A is an $n \times n$ complex matrix. Let

$$p(\lambda) = \prod_{i=1}^{k} (\lambda - c_i)^{p_i}$$

be any polynomial which annuls A, and for $1 \leq j \leq k$ put

$$f_j(\lambda) = \prod_{i \neq j} (\lambda - c_i)^{p_i}.$$

We use partial fractions to write

$$\frac{1}{p(\lambda)} = \sum_{i=1}^{k} \sum_{j=1}^{p_i} \frac{\alpha_{ij}}{(\lambda - c_i)^j},$$

and it follows that

$$\sum_{i=1}^{k} u_i(\lambda) f_i(\lambda) = 1, \tag{1.1}$$

where

$$u_i(\lambda) = \sum_{j=1}^{p_i} \alpha_{ij} (\lambda - c_i)^{p_i - j} \qquad (1 \leq i \leq k).$$

We now define, for $1 \leq j \leq k$,

$$E_j = u_j(A) f_j(A), \tag{1.2}$$

so that each E_j is a polynomial in A which can be calculated by the simple algorithm just described. These matrices have the following properties.

LEMMA 1.1. *For* $1 \leq i \leq k$, $1 \leq j \leq k$,

$$\sum_{i=1}^{k} E_i = I, \qquad E_i E_j = 0 \qquad (i \neq j), \qquad E_j^2 = E_j.$$

LEMMA 1.2. *For each* j, *the range-space* R_j *of* E_j *coincides with the null-space of* $(A - c_j I)^{P_j}$.

Simple proofs of these lemmas—using (1.1), (1.2)—will be found in [**4**]. Using the last part of Lemma 1.2, we note that $E_j \neq 0$ if, and only if, c_j is an eigenvalue of A. For this reason we assume throughout that $c_1, \ldots, c_k$ are the distinct eigenvalues of A.

We now define two further polynomials in A by

$$D = \sum_{i=1}^{k} c_i E_i, \qquad N = A - D.$$

These have the following properties.

THEOREM 1.1. *If* $A = D + N$, *where* D *and* N *are determined as above, then* D *is diagonalizable,* N *is nilpotent, and* $ND = DN$. *Moreover* D *and* N *are determined uniquely by these properties.*

A simple proof—based on the above lemmas, and on simultaneous diagonalizability for the uniqueness—is to be found in [4]. Note that the uniqueness of D implies the uniqueness of the matrices E_j. In fact, Lemma 1.1 shows that

$$D^r = \sum_{i=1}^{k} c_i^r E_i \qquad (r \geq 1),$$

and an application of Cramer's rule to the first k of these equations determines each E_j as a polynomial in D with coefficients fixed by the eigenvalues.

2. Let $m(\lambda)$ be the minimal polynomial for A, and write

$$m(\lambda) = \prod_{i=1}^{k} (\lambda - c_i)^{m_i}.$$

We now suppose that the algorithm for determining the matrices E_j, D and N is initiated by $m(\lambda)$. The following lemma plays a central role in what follows.

LEMMA 2.1. *For each j, the matrix NE_j is nilpotent of index m_j.*

Proof. By Lemma 1.1, for each j,

$$NE_j = (A - c_1 E_1 \cdots - c_k E_k) E_j = (A - c_j I) E_j.$$

and

$$(NE_j)^r = (A - c_j I)^r E_j \qquad (r \geq 1).$$

Hence,

$$(NE_j)^{m_j} = (A - c_j I)^{m_j} u_j(A) f_j(A) = u_j(A) m(A) = 0.$$

Suppose now that $(NE_j)^r = 0$. Then $(A - c_j I)^r E_j = 0$. Let $x \in \mathbb{C}^n$ and put

$$p_j(\lambda) = \prod_{i \neq j} (\lambda - c_i)^{m_i}.$$

The vector $p_j(A)x$ is in the null-space of $(A - c_j I)^{m_j}$, and so, by Lemma 1.2, in R_j. Hence, by Lemma 1.1, $p_j(A)x = E_j p_j(A)x$, and we have

$$(A - c_j I)^r p_j(A)x = (A - c_j I)^r E_j p_j(A)x = \mathbf{0}$$

for all $\mathbf{x}$ in $\mathbb{C}^n$. It follows that $(\lambda - c_j)^r p_j(\lambda)$ annuls A, and consequently $r \geq m_j$.

We can now establish the following result.

THEOREM 2.1. *If N is the nilpotent part of A, then the index of nilpotence ν for N is given by* $\nu = \max_{1 \leq j \leq k} m_j$.

From the case $\nu = 1$ we see that A is diagonalizable if, and only if, the minimal polynomial of A factors into simple linear factors. This, of course, is a well-known criterion for diagonalizability.

Proof. Lemma 1.1 shows that for $r \geq 1$

$$N^r = N^r E_1 + N^r E_2 + \cdots + N^r E_k = (NE_1)^r + (NE_2)^r + \cdots + (NE_k)^r,$$

and, hence, that $N^r = 0$ if, and only if, $(NE_j)^r = 0$ $(1 \leq j \leq k)$. Lemma 2.1 shows that this is true if, and only if, $r \geq \nu$.

We next prove

THEOREM 2.2. *If D is the diagonalizable part of A, then A and D have the same characteristic polynomial.*

Proof. Since D and N commute they are simultaneously triangulisable (see, e.g., [**6**], p. 319). A triangular matrix similar to N is nilpotent and so has zeros on its main diagonal. It follows that A and D are similar to triangular matrices with the *same* main diagonal, and so they have the same characteristic polynomial.

As a corollary we obtain a result containing the Cayley-Hamilton theorem as a special case.

COROLLARY 2.3. *The characteristic polynomial of A is a multiple of its minimal polynomial.*

Proof. As in §1, let $D = \sum_{i=1}^{k} c_i E_i$, and let R_j be the range-space of E_j. Put $d_j = \dim R_j$. Using Lemma 1.1 we obtain

$$Dx = \left(\sum_{i=1}^{k} c_i E_i \right) E_j x = c_j x$$

for any x in R_j. Now form a basis for $\mathbb{C}^n$ by first selecting a basis for each R_j and then forming their union. With respect to this basis, D is represented by $\operatorname{diag}(c_1 I_1, \ldots, c_k I_k)$, where I_i is the $d_i \times d_i$ identity matrix. Thus the characteristic polynomial of D, and so of A, is

$$\prod_{i=1}^{k} (\lambda - c_i)^{d_i}.$$

The index of nilpotence of NE_i cannot exceed d_i, and hence $m_i \leq d_i$, by Lemma 2.1. This completes the proof.

The bases for the range-spaces R_j, mentioned in the proof above, can be chosen so that their union forms a basis for $\mathbb{C}^n$ with respect to which the action of A is represented by a (similar) matrix in Jordan canonical form. The method of choice depends on the theory of nilpotent operators (see e.g. [**3**, §57]), and this theory is more subtle than anything presented here. As we shall see in §4, there are examples to which the simpler decomposition theory can be applied as an alternative to using Jordan forms. This suggests that these forms could be left as an 'optional extra' in a course which involves this theory.

3. The functional calculus. Let A be a complex $n \times n$ matrix with distinct eigenvalues c_j $(1 \leq j \leq k)$, and let D, N and the matrices E_j be constructed as in §1. Let f be a complex-valued function which has derivatives up to order $n - 1$ at each c_j, and denote by $\mathscr{F}$ the class of all such functions f. For $f \in \mathscr{F}$, we define

$$f(A) = \sum_{m=0}^{n-1} \frac{f^{(m)}(D)}{m!} N^m, \tag{3.1}$$

where, for each m,

$$f^{(m)}(D) = \sum_{i=1}^{k} f^{(m)}(c_i) E_i. \tag{3.2}$$

As far as we are aware, this definition—which seems to us particularly natural because of its connection with Taylor series—has not been given previously in this form. To make it acceptable to our students we show that it corresponds to "replacing λ by A in $f(\lambda)$" for an appropriate class of functions. For the reader, we should perhaps also show that it is equivalent to the other definitions of $f(A)$ to be found in the literature (see e.g. [**1**, chapter 5], [**2**, chapter 5], and more recently [**5**]). As a by-product, we obtain some simple results of independent interest.

THEOREM 3.1. *If $f_1, f_2 \in \mathcal{F}$, then*

$$(\lambda_1 f_1 + \lambda_2 f_2)(A) = \lambda_1 f_1(A) + \lambda_2 f_2(A),$$
$$(f_1 f_2)(A) = f_1(A) f_2(A),$$

where λ_1, λ_2 are complex scalars.

Proof. The first of these results follows immediately from (3.1) and (3.2). For the second, we use additionally the nilpotence of N and Lemma 1.1 to write

$$f_1(A)f_2(A) = \sum_{r=0}^{n-1} N^r \sum_{m=0}^{r} \frac{f_1^{(m)}(D)}{m!} \frac{f_2^{(r-m)}(D)}{(r-m)!}$$
$$= \sum_{r=0}^{n-1} \frac{N^r}{r!} \sum_{m=0}^{r} \binom{r}{m} \sum_{i=1}^{k} f_1^{(m)}(c_i) f_2^{(r-m)}(c_i) E_i.$$

Changing the order of summation, and using Leibnitz's rule, the right-hand side becomes $(f_1 f_2)(A)$.

It is easy to verify that $f(A) = A$ if $f(\lambda) = \lambda$, and that $f(A) = I$ if $f(\lambda) = 1$. It follows from this, and from Theorem 3.1, that $p(A)$ is obtained by substituting A for λ in $p(\lambda)$, whenever p is a polynomial. We can extend this to rational functions, but before doing so we first prove the following theorem.

THEOREM 3.2. *If $f \in \mathcal{F}$, then $f(D)$ is the diagonalizable part of $f(A)$, and both matrices have as their characteristic polynomial*

$$(\lambda - f(c_1))^{d_1}(\lambda - f(c_2))^{d_2} \cdots (\lambda - f(c_k))^{d_k}, \tag{3.3}$$

where $d_i = \dim R_i$.

Proof. Since $f(D) = \sum_{j=1}^{k} f(c_j) E_j$, it follows, as in the proof of Corollary 2.3, that $f(D)$ is diagonalizable, and that its characteristic polynomial is the one given in (3.3). Write

$$f(A) = f(D) + N \sum_{m=1}^{n-1} \frac{f^{(m)}(D)}{m!} N^{m-1}. \tag{3.4}$$

Since N commutes with each E_j, it follows that N commutes with each of the matrices $f^{(m)}(D)$ occurring in (3.4). So $f(A) - f(D)$ is nilpotent. Moreover this matrix commutes with $f(D)$, since each matrix is expressible as a polynomial in A. Hence $f(D)$ is the diagonalizable part of $f(A)$. An application of Theorem 2.2 completes the proof.

Now let p and q be polynomials, and assume that q is non-zero at the eigenvalues of A. Then $p/q \in \mathcal{F}$. Theorem 3.2 shows that 0 is not an eigenvalue

of $q(A)$ and so $q(A)$ is invertible. Since by Theorem 3.1 $(p/q)(A)q(A) = p(A)$ we deduce that $(p/q)(A) = p(A)(q(A))^{-1}$.

For $f \in \mathscr{F}$, it follows from (3.4) that whenever A is diagonalizable then $f(A)$ is diagonalizable. We now state a partial converse.

For a general $n \times n$ matrix M, $\sigma(M)$ will denote the set of eigenvalues of M.

THEOREM 3.3. *Let $f \in \mathscr{F}$ and suppose that $f'(\lambda) \neq 0$ whenever $\lambda \in \sigma(A)$. If $f(A)$ is diagonalizable then A is diagonalizable.*

Proof. Assume that $f(A)$ is diagonalizable. Let

$$X = \sum_{m=1}^{n-1} \frac{f^{(m)}(D)}{m!} N^{m-1},$$

where D and N are, respectively, the diagonalizable and nilpotent parts of A. We have, as in (3.4), $f(A) = f(D) + NX$, and Theorem 3.2 shows that $f(D)$ is the diagonalizable part of $f(A)$. It follows that $NX = 0$. An argument similar to the one used in the proof of Theorem 3.2 shows that $f'(D)$ is the diagonalizable part of X. Theorem 2.2 now implies that

$$\sigma(X) = \sigma(f'(D)) = \{f'(\lambda) : \lambda \in \sigma(A)\}.$$

The hypothesis shows that $0 \notin \mathrm{gs}(X)$, and so X must be invertible. We deduce that $N = 0$ and hence that A is diagonalizable.

Using Theorems 3.2 and 3.3 we can prove

THEOREM 3.4. *Let f be an entire function. Then the implication*

$$f(A) = 0 \Rightarrow A \quad \textit{diagonalizable}$$

holds for all complex $n \times n$ matrices A if, and only if, each zero of f is simple.

Proof. If each zero of f is simple then $f'(\lambda) \neq 0$ whenever $f(\lambda) = 0$. In particular if $f(A) = 0$ then $f'(\lambda) \neq 0$ whenever $\lambda \in \sigma(A)$, since for such λ, $f(\lambda) \in \sigma(f(A)) = \{0\}$. The diagonalizability of A now follows from Theorem 3.3. Suppose, conversely, that A is diagonalizable whenever $f(A) = 0$. Assume that λ_0 is a zero of f which is not simple, so that, for some entire function g, $f(\lambda) = (\lambda - \lambda_0)^2 g(\lambda)$. Let the $n \times n$ matrix A be given by $\{a_{ij}\}$, where $a_{ii} = \lambda_0$ $(1 \leq i \leq n)$, $a_{12} = 1$, and $a_{ij} = 0$ otherwise. Then $(A - \lambda_0 I)^2 = 0$, and so

$$f(A) = (A - \lambda_0 I)^2 g(A) = 0.$$

However A is not diagonalizable. This contradiction shows that each zero of f must be simple.

Since N, and each of the matrices E_j, is a polynomial in A, it is clear from (3.1) and (3.2) that $f(A)$ is a polynomial in A whenever $f \in \mathscr{F}$. Different functions f and g in $\mathscr{F}$ may therefore give rise to the same matrix. The result stated below gives a necessary and sufficient condition for this to happen, and also shows (see [5]) that alternative definitions of $f(A)$ involving Jordan canonical forms, or interpolation polynomials are equivalent to the one we give here.

THEOREM 3.5. *Let f and g be in $\mathscr{F}$. Then $f(A) = g(A)$ if, and only if,*

$$f^{(m)}(c_j) = g^{(m)}(c_j) \quad (1 \leq j \leq k, \quad 0 \leq m \leq m_j - 1), \tag{3.5}$$

where for each j, m_j is the exponent of $\lambda - c_j$ in the factorization of the minimal polynomial for A.

Proof. From (3.1), (3.2) we have

$$f(A) - g(A) = \sum_{m=0}^{n-1} \left(f^{(m)}(D) - g^{(m)}(D)\right)\frac{N^m}{m!}$$
$$= \sum_{m=0}^{n-1}\left(\sum_{j=1}^{k}\left(f^{(m)}(c_j) - g^{(m)}(c_j)\right)E_j\right)\frac{N^m}{m!}.$$

Interchanging the summations, using Lemma 2.1, and noting that $N^m E_j = (NE_j)^m$ for $m \geq 1$, the right-hand side can be written as

$$\sum_{j=1}^{k}\left(\sum_{m=0}^{m_j-1}\left(f^{(m)}(c_j) - g^{(m)}(c_j)\right)\frac{N^m E_j}{m!}\right).$$

We deduce from Lemma 1.1 that $f(A) = g(A)$ if, and only if, $(f(A) - g(A))E_j = 0$ for $1 \leq j \leq k$, and so if, and only if,

$$\sum_{m=0}^{m_j-1}\left(f^{(m)}(c_j) - g^{(m)}(c_j)\right)\frac{N^m E_j}{m!} = 0 \tag{3.6}$$

for $1 \leq j \leq k$. Hence $f(A) = g(A)$ if we have (3.5). Suppose conversely that $f(A) = g(A)$. Then the conditions (3.6) hold. Multiplying both sides of (3.6) by N^{m_j-1} shows that $f(c_j) = g(c_j)$, since by Lemma 2.1, $N^m E_j \neq 0$ for $0 \leq m \leq m_j - 1$ whereas $N^m E_j = 0$ for $m \geq m_j$. Subsequent multiplication of both sides of (3.6) by N^m for $m = m_j - 2, m_j - 3, \ldots$, successively gives the remaining equations in (3.5).

4. For some years we have given a first course in linear algebra which includes both the decomposition theory and parts of the functional calculus (as outlined here), as well as the more standard topics. We feel that the inclusion of the "new" topics provides an opportunity to set some instructive and sometimes important problems. For example a student can be asked to find the $D + N$ decomposition for a matrix A (usually 3×3 to keep things simple), and then to determine $f(A)$ for a function f (often the exponential function).

If, for example,

$$A = \begin{pmatrix} 2 & 0 & -3 \\ -1 & -1 & -3 \\ 1 & 3 & 6 \end{pmatrix},$$

then the characteristic polynomial of A is given by

$$p(\lambda) = (\lambda - 3)(\lambda - 2)^2.$$

With $c_1 = 3$, $c_2 = 2$ the algorithm of §1 now gives

$$E_1 = \begin{pmatrix} -3 & -9 & -12 \\ 0 & 0 & 0 \\ 1 & 3 & 4 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 4 & 9 & 12 \\ 0 & 1 & 0 \\ -1 & -3 & -3 \end{pmatrix},$$

so that

$$D = \begin{pmatrix} -1 & -9 & -12 \\ 0 & 2 & 0 \\ 1 & 3 & 6 \end{pmatrix}, \qquad N = \begin{pmatrix} 3 & 9 & 9 \\ -1 & -3 & -3 \\ 0 & 0 & 0 \end{pmatrix}.$$

An application of (3.1) with $f(\lambda) = e^{\lambda}$ gives

$$e^{A} = e^{3}E_1 + e^{2}E_2 + \left(e^{3}E_1 + e^{2}E_2\right)N + \tfrac{1}{2}\left(e^{3}E_1 + e^{2}E_2\right)N^{2}.$$

Rather than substitute directly for E_1, E_2 and N at this stage, we encourage students to use some theory. Thus one of the factors in $p(\lambda)$ is linear and by Lemma 2.1 this implies that $E_1N = 0$. Lemma 1.1 gives $E_1 + E_2 = I$ and hence $E_2N = N$. The highest exponent occurring in the factorization of $p(\lambda)$ is 2, and so Theorem 2.1 shows that $N^2 = 0$. This results in the simplified calculation

$$e^{A} = e^{3}E_1 + e^{2}(E_2 + N) = e^{2}\begin{pmatrix} 7 - 3e & 18 - 9e & 21 - 12e \\ -1 & -2 & -3 \\ -1 + e & -3 + 3e & -3 + 4e \end{pmatrix},$$

and illustrates a practical application of the theory.

As an illustration of the importance of the technique for finding the exponential of a square matrix, a student can be asked to solve linear differential equations of the form

$$\mathbf{y}'(t) = A\mathbf{y}(t)$$

where $\mathbf{y}(t)$ is a vector-valued function of the real variable t, and where A is a square matrix with real or complex terms. The solution, of course, involves the matrix e^{At}.

In conclusion we remark that Theorem 3.3 and Theorem 3.4 (see §3) are not included in our course work. Our knowledge of them enables us to set problems which can be both interesting and challenging. For example, Theorem 3.3 shows that "A is diagonalizable whenever e^{A} is diagonalizable." Again, Theorem 3.4 shows that "A is diagonalizable whenever $\sin A = 0$". Our students, who are not familiar with these theorems, have to use some ingenuity to prove results of this type.

REFERENCES

1. W. L. Ferrar, Finite Matrices, Clarendon Press, Oxford 1951.
2. F. R. Gantmacher, The Theory of Matrices, vol. 1, Chelsea, New York, 1959.
3. P. R. Halmos, Finite-Dimensional Vector Spaces, second edition, Van Nostrand, New York, 1958.
4. K. Hoffman and R. Kunze, Linear Algebra, second edition, Prentice-Hall, 1971.
5. Y. Ikebe and T. Inagaki, An elementary approach to the functional calculus for matrices, *American Math. Monthly*, (May 1986) 390-392.
6. L. Mirsky, An Introduction to Linear Algebra, Clarendon Press, Oxford, 1955.

Representing Binomial Coefficients as Sums of Squares

ANDREW GRANVILLE AND YILIANG ZHU, *University of Toronto*



ANDREW GRANVILLE: Having completed my B.A. and CASM at Cambridge University, I crossed the Atlantic in order to get my Ph.D. on Fermat's Last Theorem under the supervision of Paulo Ribenboim at Queen's University. Presently I am in my second year as a postdoctoral fellow at the University of Toronto. My current research interests include elementary number theory, solving Diophantine equations, studying the distributions of primes and of other special integers, and the theory of 3-designs.



YILIANG ZHU: Having completed my B.Sc. at Shanghai University of Science and Technology, I crossed the Pacific in order to pursue further studies. Life was very enjoyable during my M.Sc. Program at Queen's University. I now look forward to obtaining my Ph.D. in statistics at the University of Toronto, under the supervision of Nancy Reid.

1. Introduction. In the late eighteenth century, Lagrange proved that every positive integer can be represented as a sum of at most four squares of integers. This cannot be improved to three squares as there are infinitely many integers (e.g., 7) which cannot be represented as a sum of three squares of integers:

LEMMA 1 (Legendre, 1798). *A positive integer n cannot be expressed as the sum of three squares of integers if and only if n is of the form $n = 2^{2k}(8m + 7)$ where m and k are non-negative integers.*

In this article we shall prove two recent conjectures about whether various binomial coefficients can be written as the sum of less than four squares, using methods from the elementary theories of numbers, stochastic matrices and of graphs. First we will prove a conjecture stated by Alvan Beall, Blair Kelly, and Bob Morris at the Western Number Theory Conference in December, 1987:

THEOREM 1. *For every positive integer n, except* 1, 2, 3, 4, 5, 9, 14, 17, 18, 20, 21, 35 *and* 41, *there exists an integer $m, 0 \le m \le n$, divisible by* 4, *for which $\binom{n}{m}$ cannot be represented as a sum of three squares.*

We will also prove a conjecture of Neville Robbins [3]:

THEOREM 2. *The set of integers n, for which $\binom{2n}{n}$ cannot be represented as a sum of three squares, has asymptotic density* $1/8$ *in the set of all natural numbers.*

Our proof of Theorem 1 will rely on the following result, which we shall prove in Section 2:

PROPOSITION 1. *Suppose that m and n are positive integers with m divisible by* 4. *If $\binom{n}{m}$ cannot be written as the sum of three squares then neither can $\binom{2n}{2m}$ nor $\binom{2n+1}{2m}$.*

We define E_1 to be the set of exceptional n in the hypothesis of Theorem 1. We define S to be the set of positive integers that cannot be written as the sum of three squares of integers. We can now give the

Proof of Theorem 1. Let E_2 and E_3 be the sets of values of n that are given in Tables I and II respectively. We see there that for each such n there exists a value of m for which $\binom{n}{m} \in S$; moreover if $n \in E_3$, then m is divisible by 4.

TABLE I. The least m with $\binom{n}{m} \in S$, for each $n \in E_2$.

m	n	m	n	m	n
1	7, 28, 71, 284	9	164	17	142
2	8, 11, 43, 57, 136	10	40	18	37
3	82	11	56	19	73
5	10, 34, 83	14	36		
6	19, 165	15	68		

TABLE II. The least m, divisible by 4, with $\binom{n}{m} \in S$, for each $n \in E_3$.

n	Binary Expansion	m	n	Binary Expansion	m
16	10000	4	81	1010001	4
272	100010000	4	328	101001000	100
273	100010001	20	329	101001001	28
137	10001001	52	330	101001010	36
69	1000101	20	331	101001011	12
70	1000110	4	166	10100110	4
568	1000111000	100	167	10100111	28
569	1000111001	84	42	101010	20
285	100011101	4	86	1010110	68
143	10001111	12	87	1010111	4
72	1001000	28	22	10110	20
146	10010010	36	23	10111	4
147	10010011	44	6	110	4
74	1001010	12	112	1110000	44
75	1001011	36	113	1110001	36
38	100110	4	114	1110010	60
39	100111	12	115	1110011	4
80	1010000	20	29	11101	4
			15	1111	12

One can also see, from Table II, that for any positive integer n, not in $E_1 \cup E_2$, there exists an integer $e \in E_3$ such that the first few digits in binary notation of n are precisely those of e: in other words, there exist integers k and r such that $0 \le r \le 2^k - 1$ where

$$n = e \cdot 2^k + r.$$

(As an example, suppose that $n = 13 = (1101)_2$ in binary notation. Now $6 = (110)_2 \in E_3$ and so $13 = 6.2 + 1$. Try also $575 = (1000111111)_2$; we have $143 = (10001111)_2 \in E_3$ and so $575 = 143.2^2 + 3$.)

We shall prove that for any positive integer $n \notin E_1 \cup E_2$ there exists a value of m, divisible by 4, for which $\binom{n}{m} \in S$, by induction on k.

For $k = 0$, this is trivial as $n = e \in E_3$, and the value of m is given in Table II. So suppose the result holds for $k - 1$:

For $n = e2^k + r$ let $s = r/2$ (r even), $(r - 1)/2$ (r odd) and $p = e2^{k-1} + s$. As $0 \le s \le 2^{k-1} - 1$, we have a value of t, divisible by 4, for which $\binom{p}{t} \in S$, by the induction hypothesis. Therefore, $\binom{n}{2t} \in S$ by Proposition 1.

Remark. We see, from the proof, that if $n = e2^k + r$ where $e \in E_3$ and $\binom{e}{m}$ cannot be written as the sum of three squares, then neither can $\binom{n}{2^k m}$.

We will prove a stronger result than Theorem 2 in Section 3:

THEOREM 3. *For $a = 0, 1, 2$ and 3,*

$$\#\{n \le x : n \equiv a(\mathrm{mod}\, 4), n! \in S\} = x/32 + o(x).$$

We see that Theorem 2 follows easily from Theorem 3 by the following result, which we'll prove in Section 2:

PROPOSITION 2. *For any positive integers m and n,*

$$n \in S \quad \textit{if and only} \quad m^2 n \in S.$$

Proof of Theorem 2.

$$\begin{aligned}\#\left\{n \le x : \binom{2n}{n} \in S\right\} &= \#\{n \le x : (2n)! \in S\} \quad \text{by Proposition 2}\\ &= \#\{m \le 2x : m \equiv 0(\mathrm{mod}\, 2) \quad \text{and} \quad m! \in S\}\\ &= x/8 + o(x) \quad \text{by Theorem 3.}\end{aligned}$$

Theorem 2 is perhaps surprising, as the positive integers $n \le x$, for which $n \in S$, actually have density $1/6$. By Lemma 1,

$$\begin{aligned}\#\{n \le x : n \in S\} &= \sum_{k \ge 0} \#\{n \le x : n = 2^{2k}(8m + 7) \quad \text{for some} \quad m \ge 0\}\\ &= \sum_{k=0}^{\log x} \{x/2^{2k+3} + 0(1)\}\\ &= x/6 + 0(\log x).\end{aligned}$$

However, the reason that we get $1/8$ in Theorem 2 (instead of $1/6$) can most easily be explained by realizing that the parity of the power of 2 dividing $n!$ is equally distributed between even and odd, whereas this is not the case for ordinary integers.

2. Sequences of 0's and 1's (Proofs of Propositions 1 and 2). Any positive integer n can be written in the form $2^a b$ where a and b are nonnegative integers, and b is odd. Let $v(n)$ be the residue class of $a(\mathrm{mod}\, 2)$ and $f(n)$ be the residue class of $b(\mathrm{mod}\, 8)$. Lemma 1 may be expressed as

$$n \in S \quad \text{iff} \quad v(n) = 0 \quad \text{and} \quad f(n) = 7. \tag{1}$$

We also note that

$$v(mn) = v(m) + v(n) \quad \text{and} \quad f(mn) = f(m)f(n). \tag{2}$$

for any integers m and n. We can immediately give the

Proof of Proposition 2. By (2) we see that

$$v(m^2n) = 2v(m) + v(n) = v(n) \quad \text{and} \quad f(m^2n) = f(m)^2 f(n) = f(n).$$

Thus, by (1), $n \in S$ iff $v(n) = 0$, and $f(n) = 7$

$$\text{iff} \quad v(m^2n) = 0, \quad \text{and} \quad f(m^2n) = 7$$

$$\text{iff} \quad m^2n \in S.$$

Define $w(n) = v(n!)$ and $g(n) = f(n!)$. Also, for any fixed sequence $\alpha\beta \cdots \chi$ of 0's and 1's, define $\sigma_{\alpha\beta \ldots \chi}(n)$ to be the number of occurrences of this sequence in the binary expansion of n. We prove

LEMMA 2. *For any given positive integer* n,
(a) $w(n) \equiv n - \sigma_1(n) \pmod 2$, *and*
(b) $g(n) \equiv 3^{\sigma_{011}(n)+\sigma_{100}(n)} 7^{\sigma_{101}(n)+\sigma_{110}(n)} \pmod 8$.

Proof. Suppose that the binary expansion of n is $\sum_{i=0}^{d} a_i 2^i$ where each $a_i = 0$ or 1. Then the power of 2 dividing $n!$ is

$$\sum_{j \geq 1} \left[\frac{n}{2^j} \right] = \sum_{j \geq 1} \sum_{i=j}^{d} a_i 2^{i-j} = \sum_{i=1}^{d} a_i \sum_{j=1}^{i} 2^{i-j}$$

$$= \sum_{i=1}^{d} a_i (2^i - 1) = n - \sigma_1(n).$$

Now, as

$$\prod_{\substack{j=1 \\ j \text{ odd}}}^{m} j \equiv \begin{cases} 1 & \\ 3 & \text{if} \\ 7 & \end{cases} \begin{matrix} m \equiv 0, 1, 2, 7 \\ m \equiv 3, 4 \\ m \equiv 5, 6 \end{matrix} \pmod 8,$$

we can see that

$$g(n) = \prod_{k \geq 0} \prod_{\substack{i=1 \\ 2^k \| i}}^{n} \frac{i}{2^k} = \prod_{k \geq 0} \prod_{\substack{j=1 \\ j \text{ odd}}}^{[n/2^k]} j \pmod 8$$

$$= 3^{\#\{k \geq 0 : \tau_k = 3 \text{ or } 4\}} 7^{\#\{k \geq 0 : \tau_k = 5 \text{ or } 6\}} \pmod 8$$

where $\tau_k = 4a_{k+2} + 2a_{k+1} + a_k$, which establishes the result.

By noting that $3^a 7^b \equiv 7 \pmod 8$ if and only if a is even and b is odd, we can deduce, from Lemma 2 and (1), the following result.

COROLLARY 1. *For a given positive integer* n, $n!$ *is an element of* S *if and only if* (i) $\sigma_1(n) \equiv n \pmod 2$, (ii) $\sigma_{011}(n) \equiv \sigma_{100}(n) \pmod 2$ *and* (iii) $\sigma_{101}(n) \equiv \sigma_{110}(n) + 1 \pmod 2$.

In practice, Corollary 1 provides an efficient algorithm for determining whether $n!$ is an element of S: Simply write n in its binary notation, count the frequency of occurrences of the various digit patterns 1, 011, 100, 101 and 110, and then check whether (i), (ii), and (iii) are satisfied.

We can also deduce from Lemma 2:

COROLLARY 2. *For any positive integer* n, $w(2n+1) = w(2n) \equiv w(n) + n \pmod 2$ *and* $g(2n+1)/g(n)$ *and* $g(2n)/g(n)$ *are congruent to* 1 *and* 1, 3 *and* 1, 7 *and* 3, *and* 1 *and* 7 $\pmod 8$ *as* $n \equiv 0, 1, 2, 3 \pmod 4$, *respectively*.

Proof. $w(2n+1) - w(2n) = v(2n+1) = 0$, by definition, and $w(2n) - w(n) \equiv (2n - \sigma_1(2n)) - (n - \sigma_1(n)) \equiv n \pmod 2$ by Lemma 2(a). Also $g(2n+1)/g(2n) \equiv 2n + 1 \pmod 8$ and $g(2n)/g(n) \equiv 1, 1, 3$ and $7 \pmod 8$ as $n \equiv 0, 1, 2, 3 \pmod 4$ by Lemma 2(b).

From this, we can immediately deduce

COROLLARY 3. *Suppose that* m *is divisible by* 4 *with* $0 \le m \le n$. *Then*

$$v\left(\binom{2n+1}{2m}\right) = v\left(\binom{2n}{2m}\right) = v\left(\binom{n}{m}\right)$$

and

$$f\left(\binom{2n+1}{2m}\right) = f\left(\binom{2n}{2m}\right) = f\left(\binom{n}{m}\right).$$

Proposition 1 follows immediately from (1) and Corollary 3.

3. Proof of Theorem 3. Define $g(0) = 1, w(0) = 0$ and let

$$T(x, a; g, w) = \#\{0 \le n < x : n \equiv a \pmod 4, \quad g(n) = g \quad \text{and } w(n) = w\}$$

for

$$a = 0, 1, 2, \text{ or } 3, \quad g = 1, 3, 5 \text{ or } 7 \quad \text{and} \quad w = 0 \text{ or } 1 \ldots . \tag{3}$$

Let $p(x, a; g, w) = T(x, a; g, w)/x$, which can be thought of as the "probability" that an integer $n < x$ has the properties that $n \equiv a \pmod 4$, $g(n) = g$ and $w(n) = w$. By (1), Theorem 3 is evidently implied by

THEOREM 4. *For each* a, g *and* w *in the range* (3), *we have*

$$p(x, a; g, w) \sim 1/32 \quad \text{as} \quad x \to \infty.$$

We shall prove, in Section 4,

PROPOSITION 3. *Fix* $\varepsilon > 0$. *If* n *is sufficiently large* ($> n_\varepsilon$ *say*), *then*

$$\left| p(2^{43n}x, a; g, w) - \frac{1}{32} \right| < \varepsilon$$

for any a, g *and* w *in the range* (3) *and integer* $x \ge 1$.

From this, we can immediately give the

Proof of Theorem 4. Fix $\varepsilon > 0$ and choose $n > n_{\varepsilon/2}$ (where n_ε is as in Proposition 3). For any $z > 2^{43n}(1 + 2/\varepsilon)$ choose x to be that integer for which $2^{43n}(x-1) < z \le 2^{43n}x$. Then

$$\left| T(z, a; g, w) - \frac{z}{32} \right| \le \left| T(2^{43n}x, a; g, w) - \frac{2^{43n}x}{32} \right| + 2^{43n}$$

$$\le 2^{43n}\left(\frac{\varepsilon x}{2} + 1\right) < \varepsilon z.$$

4. Stochastic matrices. Let $p(x)$ be the 32 by 1 vector with entries $p(x, a, g, w)$. Let Q be the 32 by 32 matrix indexed in both directions by (a, g, w); where the entry in the (a', g', w')th row and (a, g, w)th column is

$$1/2 \quad \text{if} \quad a' = \begin{cases} 0 \\ 0 \\ 1 \\ 1 \\ 2 \\ 2 \\ 3 \\ 3 \end{cases} \qquad a = \begin{cases} 0 \\ 2 \\ 0 \\ 2 \\ 1 \\ 3 \\ 1 \\ 3 \end{cases} \qquad g'/g \equiv \begin{cases} 1 \\ 3 \\ 1 \\ 7 \\ 1 \\ 7 \\ 3 \\ 1 \end{cases} \pmod 8 \qquad w' - w \equiv \begin{cases} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{cases} \pmod 2$$

0 otherwise.

Note that Q is a doubly stochastic matrix all of whose entries are non-negative. (A matrix M is said to be stochastic if the sum of the entries in each of its columns is 1, doubly stochastic if the same is true for each of its rows). In Section 5 we shall prove

PROPOSITION 4. *$A = Q^{43}$ is a doubly stochastic matrix all of whose entries are positive.*

Now Corollary 2 implies that $p(2x) = Qp(x)$ for any positive integer x; by iterating this n times we get

LEMMA 3. *For any positive integers n and x, $p(2^n x) = Q^n p(x)$.*

For a given matrix M and vector x we say that $\lim_{n\to\infty} M^n x$ exists and equals y if the ith component of $M^n x$ tends to the ith component of y as n tends to ∞, for each i. An important result on stochastic matrices has been given by Perron (see [**1**, p. 216] for a nice proof).

LEMMA 4. *Suppose that M is a stochastic matrix all of whose entries are positive. There exists a vector a $(= a(M))$ such that, for* any *stochastic vector v with components all nonnegative, we have $\lim_{n\to\infty} M^n v = a$.*

We can now give a proof of Proposition 3 by using the three results directly above:

Proof of Proposition 3. $M = Q^{43}$ is a doubly stochastic matrix, all of whose entries are positive, by Proposition 4. Let v be the 32×1 vector with each entry equal to $1/32$. As each row sum of M is 1 we see that $Mv = v$, and so the value of a in Lemma 4 is given by

$$a = \operatorname*{Lim}_{n\to\infty} M^n v = \operatorname*{Lim}_{n\to\infty} v = v.$$

Therefore,

$$\operatorname*{Lim}_{n\to\infty} p(2^{43n}x) = \operatorname*{Lim}_{n\to\infty} M^n p(x) \qquad \text{by Lemma 3,}$$

$$= v \qquad \text{by Lemma 4,}$$

and the result follows immediately, by definitions.

5. Graphs of Stochastic matrices. Suppose that M is an $n \times n$ matrix, all of whose entries are non-negative. The directed graph $G(M)$ of such a matrix is obtained by taking n vertices $v_1, \dots, v_n$ and putting a directed edge from v_i to v_j if and only if the (i, j)th entry of M is non-zero. A path of length k from v to w is a sequence of k (not necessarily distinct) directed edges, starting at v and ending at w; e.g.,

$$v = a_0 \quad \text{to} \quad a_1, a_1 \quad \text{to} \quad a_2, a_2 \quad \text{to} \quad a_3, \dots, a_{k-1} \quad \text{to} \quad a_k = w.$$

From this definition we have

LEMMA 5. *Suppose that M is an $n \times n$ matrix all of whose entries are non-negative and k is a positive integer. Every entry of M^k is positive if and only if there is a directed path of length k, in both directions, between every pair of (not necessarily distinct) vertices in $G(M)$.*

Proof of Proposition 4. It is not difficult to show that the product of any two doubly stochastic matrices is itself doubly stochastic; therefore, as Q is doubly stochastic, so are $Q^2, Q^3, \dots, Q^{43} = A$.

In FIGURE 1 we give a subgraph of $G(Q)$. By visual inspection one can see that there is a directed path of length $a_v (\leq 17)$ from v to 071, and one of length $b_v (\leq 26)$ back from 071 to v, for each vertex v in $G(Q)$. Therefore, for any pair of (not necessarily distinct) vertices v and w in $G(Q)$ we let P_{vw} be the path given by joining the path of length a_v from v to 071, to $43 - a_v - b_w$ circuits of the edge from 071 to itself, to the path of length b_w from 071 to w. As P_{vw} has length 43 in each case, we know that every entry of $A = Q^{43}$ is positive, by Lemma 5.

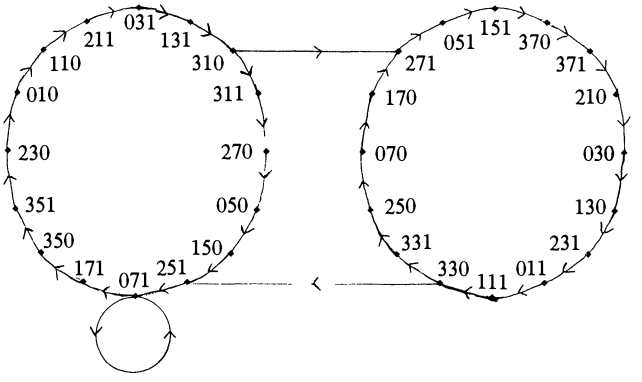


FIG. 1. The vertices are indexed by three numbers which, in order, correspond to a, g and w.

6. Some final comments. A number of related questions can be asked:

1. If we fix k what is the density of integers n for which $\binom{n}{k} \in S$? It seems possible to derive a complicated general formula that gives a variety of different values as we vary over values of k.

2. In the proof of Theorem 1 the values of $m \to \infty$ as $n \to \infty$. Is it true that there exists an integer m_0, such that for all $n \notin E_1$ we have $\binom{n}{m} \in S$ for some $m \leq m_0$? The answer to this is yes with $m_0 = 74$; however our proof is extremely

complicated and uses the idea of an *infinite* system of congruences that cover the integers (see [**2**] for a review).

3. Let $P(n)$ be the number of integers m in row n of Pascal's triangle for which $\binom{n}{m} \in S$. Let $\Pi(N) = P(1) + P(2) + \cdots + P(N)$. Does $\operatorname{Lim}_{N\to\infty} \Pi(N)/(N^2/2)$ exist? If so, what is it?, i.e., is there a density of integers in Pascal's triangle that are not representable as a sum of three squares?

REFERENCES

1. J. M. Ortega, Matrix Theory, A Second Course, Plenum, New York, 1987.
2. S. Porubsky, Results and problems on covering systems of residue classes, *Math. Seminar Giessen*, 150 (1981) 85.
3. N. Robbins, Representing $\binom{2n}{n}$ as a sum of squares, *Fib. Quarterly*, 25 (1987) 29–33.

LETTERS TO THE EDITOR

To the Editor:

In their Teaching of Mathematics note [2] M. E. Larsen and B. S. Jensen offer what they call a "simple, direct solution" for the initial-value problem of the system

$$\begin{aligned} \dot{x}_1 &= ax_1 + bx_2 \\ \dot{x}_2 &= cx_1 + dx_2 \end{aligned} \tag{1}$$

which avoids transformation to coordinates along the eigenvectors of the coefficient matrix A and the use of complex eigenvalues. In my opinion, their solution is neither simple nor direct and, worst of all, from a pedagogical point of view, in spite of its elaborate methodology, the solution applies only to 2×2 systems, where the characteristic equation for A is simply $A^2 = -\det(A)I$ if $\operatorname{tr}(A) = 0$.

The Gaussian elimination method, which applies to the most general systems (see, e.g. [1, Chap. 6]) gives, even for the special system (1), a simpler and more direct solution, without the use of eigenvalues and other results from linear algebra. Write (1) as

$$(D - a)x_1 - bx_2 = 0, \quad -cx_1 + (D - d)x_2 = 0$$

and eliminate x_2 by applying the operators $(D - d), b$, obtaining

$$\left[(D - \Theta)^2 - \Delta\right]x_1 = 0, \tag{2}$$

where we have used the notations of [2]: $\Theta = (a + d)/2\Delta = ((a - d)/2)^2 + bc$. Using the shifting rule, according to which $p(D + h)x = e^{-ht}p(D)e^{ht}x$, (2) is equivalent to

$$(D^2 - \Delta)e^{-\Theta t}x_1 = 0. \tag{3}$$

From here on, or even immediately after (2), one could refer to the universally taught method for solving second-order linear equations with constant coefficients; or proceeding as in (2), let $y = \delta$ be the solution of $(D^2 - \Delta)y = 0$ for which $y(0) = 0$, $\dot{y}(0) = 1$. Then $y = \dot{\delta}$ is the solution for which $y(0) = 1$, $\dot{y}(0) = 0$, and the general solution of (3) is

$$x_1(t) = e^{\Theta t}\left(c_1\delta(t) + c_2\dot{\delta}(t)\right), \tag{4}$$

hence $x_1(0) = c_2$, $\dot{x}_1(0) = c_1 + \Theta c_2$. The constants c_1, c_2 are determined from the given values $x_1(0) = x_1^0$, $x_2(0) = x_2^0$, thus by (1) $\dot{x}_1(0) = ax_1^0 + bx_2^0$. Then (4) gives

$$x_1(t) = e^{\frac{a+d}{2}t}\left(\left(\frac{a - d}{2}x_1^0 + bx_2^0\right)\delta + x_1^0\dot{\delta}\right),$$

which is the solution of [2]. By interchanging a and d, b and c, x_1^0 and x_2^0, one obtains x_2.

MICHAEL GOLOMB
Purdue University

REFERENCES

1. M. Golomb and M. E. Shanks, Elements of Differential Equations, 2nd ed., McGraw Hill, New York, 1965.
2. M. E. Larsen and B. S. Jenson, Coupled linear differential equations with real coefficients, this MONTHLY, 96 (1989) 727–731.

To the Editor:

The result in the note, "Counting the Rationals" by Y. Sagher in the November MONTHLY appeared in your journal 29 years ago. See K. McCrimmon, Enumeration of the positive rationals, *Amer. Math. Monthly*, 67 (1960) 868.

Yours sincerely,

HARVEY I. BLAU
Professor of Mathematical Sciences
Northern Illinois University
DeKalb, Illinois 60115

To the Editor:

In my review of *Complex Analysis* by Bak and Newman and *Invitation to Complex Analysis* by Boas (March, 1990), I noted my suprise not to find the symmetric version of Rouché's theorem ("if f and g are analytic on and inside the simple closed curve Γ and $|f+g| < |f| + |g|$ on Γ, then f and g have the same number of zeroes inside Γ") in either book and added "true, no other text of my acquaintance includes it either; but somehow one comes to expect more from Boas and Newman." In fact, this result does appear as Theorem 18.5 on page 156 of the text **[E]. An undoubtedly more accessible reference for readers of the *Monthly* is [G].**

Sincerely yours,

LAWRENCE ZALCMAN
Bar-Ilan University

[E] T. Estermann, Complex Numbers and Functions, Athlone Press, London, 1952.
[G] Irving Glicksberg, A remark on Rouché's theorem, this *MONTHLY*, 83 (1976) 186–187.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

A Geometric Regularity From Order-Restricted Statistical Inference*

TIM ROBERTSON and CAROLYN PILLERS
Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242

In the course of our investigations in statistical inference, we discovered that the following identity is true for all integers $k \geqslant 3$:

$$\binom{k-1}{2}^{-1} \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \cos^{-1} \sqrt{\frac{i(k-j)}{(k-i)j}} = 60°,$$

where $\cos^{-1}$ is the principal value of the inverse cosine function. This is an unusual geometric identity because it deals with averages of angles. We will first prove the identity, and then sketch its geometric interpretation and connection to statistical inference.

Proof. First note that $\binom{k-1}{2}$ is one more than a multiple of three when k is a multiple of three, and is a multiple of three otherwise. Let S be the set of $\binom{k-1}{2}$ indices, (i, j), involved in the double sum appearing in the identity. A graph of S, which is a triangular-shaped lattice in R^2, is given in Figure 1 for $k = 9$. Partition S into groups of three, with possibly one element remaining, and enumerate the points in the following manner: The corners of the triangle are the first three points. The next set of three points consists of the point just below the upper left corner together with the highest and lowest points of the vertical line just to the left of the upper right corner. Each succeeding set of three points consists of one

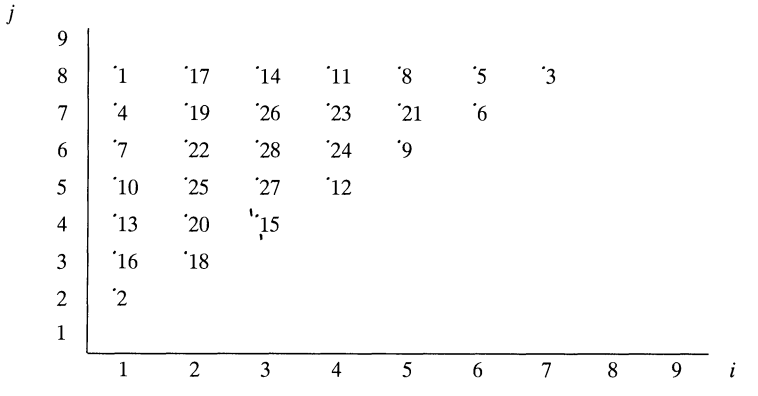


FIG. 1

*This research was sponsored by the Office of Naval Research under ONR contract N00014-80-C0321.

point from each edge of the triangle as in FIGURE 1. When the outside triangle is exhausted by this process, the enumeration is continued on the next, inner, triangle. The proof consists of showing that the sum of the angles corresponding to each set of three points is 180°. If k is divisible by three so that there is a point left over at the end, then the corresponding angle is exactly 60°.

A triple which corresponds to the corners of one of the triangles is of the form $(m, k-m), (m, 2m), (k-2m, k-m)$ for some positive integer m. Thus, the corresponding angles are

$$\cos^{-1}\sqrt{m^2/[(k-m)(k-m)]}\,, \quad \cos^{-1}\sqrt{[m(k-2m)]/[2m(k-m)]}\,,$$

and

$$\cos^{-1}\sqrt{[(k-2m)m]/[2m(k-m)]}\,.$$

By combining the last two angles using the inverse version of the double angle formula, $2\cos^{-1}x = \cos^{-1}(2x^2-1)$, it follows that these three angles add up to 180°.

Any other triple must be of the form $(m, k-m-n), (k-2m-n, k-m)$, $(k-2m-n, k-m-n)$ for positive integers m and n. By combining two of the associated angles using the inverse version of the angle sum formula, $\cos^{-1}x + \cos^{-1}y = \cos^{-1}[xy - (1-x^2)^{1/2}(1-y^2)^{1/2}]$, it follows that these three angles add up to 180°.

If there is a point remaining, then $k = 3a$ for some positive integer a and the remaining point is $(a, 2a)$. The associated angle has cosine equal to $\sqrt{[a(k-2a)]/[2a(k-a)]} = \sqrt{a^2/(2a)^2}$ and must be 60°.

Thus, the double sum appearing in the identity is $180° \times \binom{k-1}{2}/3$ if $\binom{k-1}{2}$ is a multiple of 3, and is $180° \times \left[\binom{k-1}{2}/3\right] + 60°$ if $\binom{k-1}{2}$ is one more than a multiple of 3. Since there are $\binom{k-1}{2}$ terms in the double sum, the average measure must be exactly 60°.

A referee suggested an alternative proof using the following lemma: Let a, b, c be nonnegative with $a+b>0, b+c>0, a+c>0$. Let $\theta_1, \theta_2, \theta_3$ be acute angles defined by $\cos^2\theta_1 = bc/[(a+b)(a+c)]$, $\cos^2\theta_2 = ac/[(a+b)(b+c)]$, $\cos^2\theta_3 = ab/[(a+c)(b+c)]$. Then $\theta_1+\theta_2+\theta_3 = 180°$. The theorem then follows by setting $a = k-2m-n$, $b = m+n$, $c = m$.

Remarks. This identity states that the average of the angles between the generators of a particular polyhedral cone in Euclidean k-space is 60° (see Stoer and Witzgall [**4**] for a thorough treatment of polyhedral cones). We discovered this identity as part of our research in an area of statistics called order restricted inference. An account of the theory in this area can be found in [**3**] and more concise summaries in [**1**], [**2**], and [**5**].

Polyhedral cones play important roles in order restricted inference as restricted estimates are found by projecting the usual unrestricted estimates onto the cone and test statistics are proportional to the distances between the various estimates. One cone that is of particular interest is $L_k = \{\mathbf{x};\ x_1 \leqslant x_2 \leqslant \cdots \leqslant x_k\}$. This cone is the direct sum of the pointed cone, $\tilde{L}_k = L_k \cap \{\mathbf{x};\ \Sigma_{i=1}^k x_i = 0\}$ and the linear subspace, $\{\mathbf{x};\ x_1 = x_2 = \cdots = x_k\}$ and it is the pointed cone $\tilde{L}_k$ that has the

regularity expressed in the identity. The generators of $\tilde{L}_k$ are given by

$$\begin{aligned}
\mathbf{l}_1 &= (-k+1, 1, 1, \ldots, 1) \\
\mathbf{l}_2 &= (-k+2, -k+2, 2, 2, \ldots, 2) \\
&\vdots \\
\mathbf{l}_{k-1} &= (-1, -1, \ldots, -1, k-1).
\end{aligned}$$

For $i < j$, the cosine of the angle between $\mathbf{l}_i$ and $\mathbf{l}_j$ is equal to $\sqrt{[i(k-j)]/[(k-i)j]}$. There are $\binom{k-1}{2}$ distinct pairs of generators and the left side of the identity represents the average of the angles between the generators.

REFERENCES

1. D. J. Bartholomew, Isotonic Inference, Encyclopedia of Statistical Sciences, Vol. 4, John Wiley and Sons, New York, 1983, pp. 260–265.
2. Tim Robertson, Monotone Relationships, Encyclopedia of Statistical Sciences, Vol. 5, John Wiley and Sons, New York, 1985, pp. 609–611.
3. Tim Robertson, F. T. Wright, and R. L. Dykstra, Order Restricted Statistical Inference, John Wiley and Sons, London, 1988.
4. J. Stoer and C. Witzgall, Convexity and Optimization in Finite Dimensions I, Springer-Verlag, Berlin, 1970.
5. F. T. Wright, Order-restricted inferences, Encyclopedia of Statistical Sciences, Vol. 6, John Wiley and Sons, New York, 1985, pp. 501–503.

The Solution of Certain Integral Equations by Means of Operators of Arbitrary Order

BERTRAM ROSS and BALDEV K. SACHDEVA
University of New Haven, West Haven, CT 06516

An integral of the form

$$\frac{1}{\Gamma(\nu)} \int_c^x (x-t)^{\nu-1} f(t)\, dt, \qquad \mathrm{Re}(\nu) \geqq 0 \tag{1}$$

is known as the Riemann-Liouville integral. It defines integration and differentiation to an arbitrary order. The R-L integral can be denoted by ${}_cD_x^{-\nu} f(x)$. The subscripts on the operator D denote the terminals of integration. The fractional calculus is not limit derived; hence, the words "terminals of integration" are more appropriate than "limits of integration". When $\nu = n$ an integer, the operators ${}_oD_x^{-\nu}$ and ${}_oD_x^{\nu}$ are, respectively, ordinary integration and differentiation.

The idea of differentiation to an arbitrary order started in 1695 when L'Hôpital asked Leibniz what would happen with $d^n y/dx^n$ when $n = 1/2$. Subsequently, the topic started with the misnomer fractional calculus because ν can be rational, irrational or complex. After a half century of controversy due to the lack of a precise definition of fractional derivatives, the matter was finally settled in 1886 by Laurent who developed (1) starting with Cauchy's integral formula in the complex plane.

One of the applications of the fractional calculus is the simplification of the solution of certain integral equations as shown in [1]. Abel, in 1823, was the first to apply the fractional calculus in the solution of the integral equation

$$\int_0^x (x-t)^{\nu-1} f(t)\,dt = k,$$

where k is a constant, which arises in the formulation of the tautochrone (isochrone) problem and other physical problems [2].

The solution of the integral equation

$$f(x) + \lambda \int_c^x K(x,t) f(t)\,dt = g(x)$$

was originally attained by Volterra in 1896. When $K(x,t)$ is a difference kernel of the form $K(x-t)$, the standard technique for solving the above when $c = 0$ is the use of Laplace transforms provided that the Laplace transform (one-sided) of the functions $f(x)$, $K(x)$, and $g(x)$ exist. Volterra's method which stemmed from the work of J. C. F. Sturm (1836) was attained as the limiting form of a set of algebraic equations in which differences of equally spaced points on (c, x) were used.

This paper has a dual purpose. One purpose is to exemplify the power and elegance of the Riemann-Liouville operators in the solution of a Volterra type equation of the form

$$f(x) + \frac{1}{\Gamma(\nu)} \int_0^x (x-t)^{\nu-1} f(t)\,dt = g(x), \tag{2}$$

where $g(x)$ is a known function and where $f(0)$ and $f'(0)$ are assumed to be finite. The exponent on the kernel can be arbitrary with the exception that $\nu - 1 \neq$ a negative integer.

A second purpose of this note is to introduce a method of solving (2) which is conceptually different than that of Volterra, Fredholm, and Laplace transforms. This method consists of repeatedly applying R-L operators of appropriate order until (2) is transformed into an ordinary differential equation. This transformed equation is called the rational equivalent. The nonhomogeneous term $g(x)$, as a result of repeated operations, gives rise to generalized integrals and derivatives. Thus, $g(x)$ is in a wide class of functions such that $D^{\pm q} g(x)$ is calculable. A table of integrals and derivatives to an arbitrary order is given in [3]. As an example to illustrate the procedure, we will specify $\nu = 2/3$ and $g(x) = x^2$.

Eq. (2) is written in operator form:

$$f(x) + {}_0D_x^{-2/3} f(x) = x^2. \tag{3}$$

Operate on (3) with $D^{-1/3}$ and $D^{1/3}$. Omitting subscripts on D for convenience, we get the two equations

$$D^{-1/3} f(x) + D^{-1} f(x) = D^{-1/3} x^2 \tag{4}$$

and

$$D^{1/3} f(x) + D^{-1/3} f(x) = D^{1/3} x^2. \tag{5}$$

Operate on (4) with $D^{4/3}$ which yields

$$Df(x) + D^{1/3} f(x) = Dx^2. \tag{6}$$

In (6) substitute for $D^{1/3}f(x)$ which is obtained from (5) getting

$$Df(x) - D^{-1/3}f(x) + D^{1/3}x^2 = Dx^2. \tag{7}$$

In (7) substitute for $D^{-1/3}f(x)$ from Eq. (4) and we will have

$$Df(x) - \left(-D^{-1}f(x) + D^{-1/3}x^2\right) + D^{1/3}x^2 = Dx^2, \tag{8}$$

or

$$Df(x) + D^{-1}f(x) = Dx^2 - D^{1/3}x^2 + D^{-1/3}x^2. \tag{9}$$

Finally, operate on each term above with D:

$$D^2f(x) + f(x) = D^2x^2 - D^{4/3}x^2 + D^{2/3}x^2$$

which we write as

$$f''(x) + f(x) = 2 - D^{4/3}x^2 + D^{2/3}x^2. \tag{10}$$

In developing Eq. (10) which is the rational equivalent of (2), it is important to note the use of the index laws for integration and differentiation to an arbitrary order. In the Riemann-Liouville fractional calculus the index law for integration to an arbitrary order

$${}_oD_x^{-p}({}_oD_x^{-q}f(x)) = {}_oD_x^{-(p+q)}f(x)$$

generally holds true; however, the index law for differentiation to an arbitrary order

$${}_oD_x^{p}({}_oD_x^{q}f(x)) = {}_oD_x^{p+q}f(x)$$

holds only for certain p, q, and $f(x)$, [2]. This is one of the subtle perils in dealing with fractional operators. It is worth examining the operation we did earlier with $D^{4/3}$. We assumed $D^{4/3}(D^{-1/3}x^2) = Dx^2 = 2x$.

The above result can be readily verified by first computing ${}_oD_x^{-1/3}x^2$. From the list of formulas for integration and differentiation to an arbitrary order [3], we have the formula for integration to an arbitrary order

$${}_oD_x^{-\nu}x^a = \frac{\Gamma(a+1)}{\Gamma(a+\nu+1)}x^{a+\nu} \qquad \mathrm{Re}(\nu) \geqq 0, \qquad a \geqq 0.$$

$${}_oD_x^{-1/3}x^2 = \frac{\Gamma(3)}{\Gamma(10/3)}x^{7/3}.$$

$${}_oD_x^{-1/3}x^2 = \frac{\Gamma(3)}{\Gamma(10/3)}x^{7/3}.$$

For differentiation to an arbitrary order we have the formula

$${}_oD_x^{\nu}x^a = \frac{\Gamma(a+1)}{\Gamma(a-\nu+1)}x^{a-\nu}, \qquad \mathrm{Re}(\nu) \geqq 0, \qquad a \geqq 0.$$

$${}_oD_x^{4/3}\left(\frac{\Gamma(3)}{\Gamma(10/3)}x^{7/3}\right) = \frac{\Gamma(3)}{\Gamma(10/3)} \cdot \frac{\Gamma(10/3)}{\Gamma((7/3)-(4/3)+1)}x = 2x. \tag{11}$$

We compute the second and third terms of the right side of (10) with the above formula (11).

$$D^{4/3}x^2 = \frac{2}{\Gamma(5/3)}x^{2/3} = K_1x^{2/3} \tag{12}$$

$$D^{2/3}x^2 = \frac{2}{\Gamma(7/3)}x^{4/3} = K_2x^{4/3}. \tag{13}$$

The differential equation (10) can now be written as

$$y'' + y = 2 - K_1x^{2/3} + K_2x^{4/3}. \tag{14}$$

The solution to (14) is the complementary solution

$$c_1 \sin x + c_2 \cos x$$

plus the particular solutions. The particular solution for 2 is clearly 2. To obtain y_p for $-K_1x^{2/3}$ the method of variation of parameters is used. The Wronskian

$$W(\sin x, \cos x) = \begin{vmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{vmatrix} = -1.$$

Identifying $y_1 = \sin x$ and $y_2 = \cos x$, we have

$$u_1' = \frac{-y_2f(x)}{W} = \frac{-(\cos x)(-K_1x^{2/3})}{-1} = -K_1x^{2/3}\cos x,$$

and

$$u_1 = -K_1\int_0^x x^{2/3}\cos x\,dx.$$

$$u_2' = \frac{y_1f(x)}{W} = \frac{(\sin x)(-K_1x^{2/3})}{-1} = K_1x^{2/3}\sin x,$$

and

$$u_2 = K_1\int_0^x x^{2/3}\sin x\,dx.$$

Thus, the particular solution for $-K_1x^{2/3}$ is

$$-K_1 \sin x \int_0^x x^{2/3}\cos x\,dx + K_1 \cos x \int_0^x x^{2/3}\sin x\,dx.$$

Similarly, we find the particular solution for $K_2x^{4/3}$ to be

$$K_2 \sin x \int_0^x x^{4/3}\cos x\,dx - K_2 \cos x \int_0^x x^{4/3}\sin x\,dx.$$

The solution to (14) is thus

$$y = f(x) = c_1 \sin x + c_2 \cos x + 2 + \sin x \int_0^x (K_2x^{4/3} - K_1x^{2/3})\cos x\,dx$$

$$+ \cos x \int_0^x (K_1x^{2/3} - K_2x^{4/3})\sin x\,dx. \tag{15}$$

We can determine the constants c_1 and c_2 from the original integral equation which we rewrite for convenience:

$$f(x) + \frac{1}{\Gamma(2/3)}\int_0^x (x-t)^{-1/3} f(t)\,dt = x^2. \tag{16}$$

We note in the above that $f(0) = 0$. Thus, we take $c_2 = -2$ to meet this condition. Thus, the solution is

$$f(x) = c_1 \sin x - 2\cos x + 2 + \sin x \int_0^x \cdots + \cos x \int_0^x \cdots . \tag{17}$$

To determine c_1 we take the derivative of the original integral equation (16) and determine $f'(0)$:

$$f'(x) + \frac{d}{dx}\frac{1}{\Gamma(2/3)}\int_0^x (x-t)^{-1/3} f(t)\,dt = 2x. \tag{18}$$

Leibniz's rule for differentiating an integral with respect to a parameter is

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(t,x)\,dt = \int_{a(x)}^{b(x)} \frac{\partial f(t,x)\,dt}{\partial x} + f[b(x),x]b'(x) - f[a(x),x]a'(x).$$

The derivative of the integral in (18) is then

$$-\frac{1}{3}\int_0^x (x-t)^{-4/3} f(x)\,dx + f[x,x](1) - 0.$$

For $f'(0)$ Eq. (18) is

$$f'(0) = 0 + \left[\lim_{t\to x} \frac{f(t)}{(x-t)^{1/3}}\right]_{x=0} - 0. \tag{19}$$

Passing to the limit and then letting $x = 0$ yields the indeterminate form $0/0$ because $f(0) = 0$. The application of L'Hôpital's rule yields

$$\lim_{x\to 0}\left[\lim_{t\to x} -3(x-t)^{2/3} f'(t)\right]$$

obtaining

$$f'(0) = 0 \cdot f'(0) = 0. \tag{20}$$

To determine the constant c_1 in (17), we take the derivative of (17) and set $x = 0$ which gives

$$f'(0) = c_1 \cos 0.$$

Thus $c_1 = 0$ and the solution is

$$f(x) = 2 - 2\cos x + \sin x \int_0^x \left(K_2 x^{4/3} - K_1 x^{2/3}\right)\cos x\,dx + \cos x \int_0^x \left(K_1 x^{2/3} - K_2 x^{4/3}\right)\sin x\,dx, \tag{21}$$

where $K_1 = 2/\Gamma(5/3)$ and $K_2 = 2/\Gamma(7/3)$.

To get a good feel for the efficiency of the method of fractional operators, we can compare it to the method of Laplace transforms. Our original integral equation is

$$f(x) + \frac{1}{\Gamma(2/3)}\int_0^x (x-t)^{-1/3} f(t)\,dt = x^2.$$

From the theory of convolution integrals we can write

$$L\{f(x)\} + \frac{1}{\Gamma(2/3)} L\{x^{-1/3} * f(x)\} = L\{x^2\},$$

$$F(s) + \frac{1}{\Gamma(2/3)} L\{x^{-1/3}\} \cdot L\{f(x)\} = \frac{2}{s^3}.$$

Then

$$F(s) + \frac{1}{\Gamma(2/3)} \cdot \frac{\Gamma(2/3)}{s^{2/3}} F(s) = \frac{2}{s^3},$$

$$F(s) = \frac{2}{s^3} \cdot \frac{s^{2/3}}{s^{2/3}+1} = \frac{2}{s^{7/3}(s^{2/3}+1)}.$$

The inverse Laplace transform of $F(s)$ is $f(x)$. However, the inverse of the function on the right above requires a substantial effort to obtain. This effort supports the assertion that the method of fractional operators is indeed an efficient one.

REFERENCES

1. Bertram Ross, Serendipity in mathematics, *American Mathematical Monthly*, (October 1983) 562.
2. K. Oldham and J. Spanier, Fractional Calculus, Academic Press, NY, 1974.
3. Bertram Ross, Methods of Summation, Descartes Press, Koriyama, Japan, 1987, p. 84.

Relations Between $\omega(n)$ and $\Omega(n)$

ARMEL MERCIER
Département de mathématiques, Université du Québec à Chicoutimi, Chicoutimi, QC, Canada, G7H 2B1

For a positive integer n, let $\omega(n)$ and $\Omega(n)$ denote, respectively, the number of distinct prime factors and the total number of prime factors of n. That is, $\omega(1) = \Omega(1) = 0$ and if $n = p_1^{e_1} \cdots p_r^{e_r}$ then $\omega(n) = r$ and $\Omega(n) = e_1 + \cdots + e_r$. Relations between $\omega(n)$ and $\Omega(n)$ are not easy to obtain directly from their definitions. However, many relations can be deduced with the help of Euler products. For example, we will show that for every real t we have the identity

$$t^{\omega(mn)-\omega(m)} = \sum_{d|(m,n)} t^{\omega(n/d)}(1-t)^{\Omega(d)}, \tag{1}$$

where (m, n) is the gcd of m and n.

To prove (1), let $m = p_1^{a_1} \cdots p_r^{a_r}$, $a_j \geqslant 1$, the p_j being distinct prime numbers. Then for each positive integer n and each j $(1 \leqslant j \leqslant r)$ there exists a unique integer i_j such that

$$n = p_1^{i_1} \cdots p_r^{i_r} M, \quad (M, m) = 1.$$

If f is multiplicative and if $\Sigma f(n)$ is absolutely convergent, then by Theorem 11.6 of [1], we have

$$\begin{aligned}
\sum_{n=1}^{\infty} f(mn) &= \sum_{\substack{i_j \geqslant 0 \\ (m, M)=1}} f\left(p_1^{a_1+i_1} \cdots p_r^{a_r+i_r} M\right) \\
&= \prod_{p^a \| m} \left(f(p^a) + f(p^{a+1}) + \cdots\right) \sum_{\substack{M=1 \\ (M, m)=1}}^{\infty} f(M) \\
&= \prod_{p^a \| m} \left(\frac{f(p^a) + f(p^{a+1}) + \cdots}{1 + f(p) + f(p^2) + \cdots}\right) \sum_{n=1}^{\infty} f(n)
\end{aligned}$$

where $p^a \| m$ means $p^a | m$ and $p^{a+1} \nmid m$, provided no factor $1 + f(p) + f(p^2) + \cdots$ vanishes.

Consequently, when $f(n) = t^{\omega(n)} n^{-s}$ and $\mathscr{R}s > 1$ ($\mathscr{R}s$ is the real part of s), we can write

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{t^{\omega(mn)}}{n^s} &= \sum_{k=1}^{\infty} \frac{t^{\omega(k)}}{k^s} t^{\omega(m)} \prod_{p | m} \left(\frac{1 + \dfrac{1}{p^s} + \dfrac{1}{p^{2s}} + \cdots}{1 + \dfrac{t}{p^s} + \dfrac{t}{p^{2s}} + \cdots}\right) \\
&= \sum_{k=1}^{\infty} \frac{t^{\omega(k)}}{k^s} t^{\omega(m)} \prod_{p | m} \left(\frac{1}{1 + \dfrac{t-1}{p^s}}\right) \\
&= \sum_{k=1}^{\infty} \frac{t^{\omega(k)}}{k^s} t^{\omega(m)} \sum_{d | m} \frac{(1-t)^{\Omega(d)}}{d^s}.
\end{aligned}$$

Now we put $kd = n$ and equate coefficients of n^{-s} to obtain

$$t^{\omega(mn)} = t^{\omega(m)} \sum_{d | (m, n)} t^{\omega(n/d)} (1-t)^{\Omega(d)},$$

which is equivalent to (1).

One can obtain other interesting nonobvious relations from (1). For example if we put $m = n$, we obtain

$$1 = \sum_{d | n} t^{\omega(n/d)} (1-t)^{\Omega(d)}, \quad \text{for all } t,$$

and differentiating termwise we get another relation.

Using the Euler product we can get many other identities. For example, for any positive integer m and for $|t| < 2^{\mathscr{R}s}$, $\mathscr{R}s > 1$, we easily obtain for all t,

$$\sum_{d|n} (1-t)^{\Omega(dm)} t^{\omega((n/d)m)} = (1-t)^{\Omega(m)} t^{\omega(m)} \sum_{d|(m,n)} (1-t)^{\Omega(d)},$$

if we consider the expansion of the product

$$\sum_{n=1}^{\infty} \frac{t^{\omega(mn)}}{n^s} \sum_{n=1}^{\infty} \frac{(1-t)^{\Omega(mn)}}{n^s}.$$

Acknowledgements. The author would like to thank the referee for a number of suggestions that improved the final version of this note.

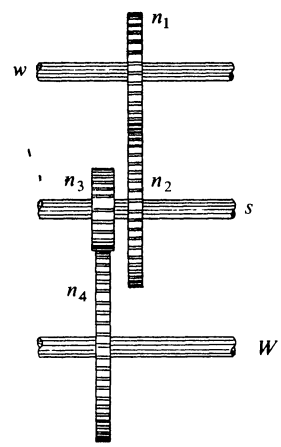
REFERENCE

1. Tom M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, 1976.

Gear Trains and Continued Fractions

JEROLD MATHEWS
Department of Mathematics, Iowa State University, Ames, IA 50011

Introduction. Gears have been used to change the rate or direction of rotary motion for at least 5,000 years. In the 27th century B.C. the Chinese used a gear train to connect the steering mechanism of a chariot to the base of a mounted statuette with outstretched arm, so that the arm would remain south-pointing as the chariot moved. A common gear train is shown in the figure, where an input axle with angular speed of w revolutions per minute (rpm) drives a gear with n_1 teeth, which drives a second gear with n_2 teeth, . . . , which, finally, drives an output axle at W rpm.



The input axle drives the first gear at $n_1 w$ teeth/minute, which is transmitted to the second gear. Letting the angular speed of the intermediate axle be s, we have $n_2 s = n_1 w$. The third gear is driven at $n_3 s$ teeth/minute and the fourth at $n_3 s = n_4 W$. From these relations we have $n_4 W = n_3 s = n_3(n_1 w/n_2)$, or

$$\frac{W}{w} = \frac{n_1 n_3}{n_2 n_4}. \tag{1}$$

If the number n of teeth on a gear must be between 20 and 120, then to find a gear train that transforms an input rpm w to an output rpm W as closely as possible we must

$$\begin{aligned} &\text{Find } n_1, n_2, n_3, n_4 \in \{20, \ldots, 120\} \\ &\text{so that } \left| \frac{W}{w} - \frac{n_1 n_3}{n_2 n_4} \right| \text{ is a minimum.} \end{aligned} \tag{2}$$

If $n_1, n_2, n_3, n_4 \in \{20, \ldots, 120\}$, then $x = (n_1 n_3)/(n_2 n_4)$ is a *feasible* solution. A solution of (2) is called *optimal*. Since there are at most 101^4 feasible solutions, (2) has an optimal solution. The bounds 20 and 120 give practicable limits on the number of teeth that can be cut on a gear blank.

Piper [**6**] gave in 1953 an algorithm for solving (2). His method is based upon elementary properties of continued fractions. More recently Orthwein [**4**] presented a somewhat different algorithm, also based on continued fractions. His algorithm does not appear to give an optimum ratio in all cases. We describe Piper's algorithm here, using notation and results given in Hardy and Wright's book *An Introduction to the Theory of Numbers* [**1**].

Simple Continued Fractions. If we divide 826 by 773 we find that

$$\frac{826}{773} = 1 + \frac{53}{773} = 1 + \cfrac{1}{\cfrac{773}{53}}. \tag{3}$$

Continuing the division, $773/53 = 14 + 31/53$ and hence

$$\frac{826}{773} = 1 + \cfrac{1}{14 + \cfrac{1}{\cfrac{53}{31}}}. \tag{4}$$

Repeating this process until a remainder of 0 is found, we obtain

$$\frac{826}{773} = 1 + \cfrac{1}{14 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{4}}}}}}. \tag{5}$$

If at each step we compute the value of the fraction up through the most recently obtained quotient we find

$$1, 1 + 1/14, 1 + 1/(14 + 1/1), 1 + 1/(14 + 1/(1 + 1/1)), \ldots$$

or

$$\frac{1}{1}, \frac{15}{14}, \frac{16}{15}, \frac{31}{29}, \frac{78}{73}, \frac{187}{175}, \frac{826}{773}. \tag{6}$$

These numbers are called *convergents* to the continued fraction (5). The quotients $1, 14, 1, 1, 2, 2, 4$ are called *partial quotients*.

More generally, if x is any positive rational we may obtain a continued fraction equal to x through the division algorithm:

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots + \cfrac{1}{a_{N-1} + \cfrac{1}{a_N}}}}}. \tag{7}$$

The partial quotients are $a_0, a_1, \ldots, a_N$; the convergents are $p_0/q_0 = a_0$, $p_1/q_1 = (a_1a_0 + 1)/a_1, \ldots, p_N/q_N = x$. The continued fraction (7) is often written as $[a_0, a_1, a_2, \ldots, a_N]$.

THEOREM A. *The convergents $p_0/q_0, p_1/q_1, \ldots, p_N/q_N$ are related to the partial quotients $a_0, a_1, \ldots, a_N$ through the relations*

$$p_0 = a_0,\ p_1 = a_1a_0 + 1,\ p_n = a_np_{n-1} + p_{n-2}, \qquad 2 \leqslant n \leqslant N, \tag{8}$$

$$q_0 = 1,\ q_1 = a_1, \qquad q_n = a_nq_{n-1} + q_{n-2}, \qquad 2 \leqslant n \leqslant N, \tag{9}$$

and

$$[a_0, a_1, a_2, \ldots, a_n] = p_n/q_n \qquad 0 \leqslant n \leqslant N \tag{10}$$

THEOREM B. *The convergents and partial quotients satisfy*

$$x = [a_0, a_1, a_2, \ldots, a_N] = [a_0, a_1, a_2, \ldots, a_{n-1}, [a_n, a_{n+1}, \ldots, a_N]] \tag{11}$$

and

$$p_nq_{n-1} - p_{n-1}q_n = (-1)^{n-1} \quad \textit{for} \quad 1 \leqslant n \leqslant N, \tag{12}$$

and

$$x = \frac{[a_n, a_{n+1}, \ldots, a_N]p_{n-1} + p_{n-2}}{[a_n, a_{n+1}, \ldots, a_N]q_{n-1} + q_{n-2}} \quad \textit{for} \quad 2 \leqslant n \leqslant N. \tag{13}$$

Letting $x_n = p_n/q_n$ denote the nth convergent, then

$$x_0 \leqslant x_2 \leqslant \cdots \leqslant x \leqslant \cdots \leqslant x_3 \leqslant x_1. \tag{14}$$

All of the inequalities in (14) are strict except the one between x and the last convergent, which is an equality. The convergents to x have no common factors.

THEOREM C. *If* $N > 1$ *and* $1 \leqslant n \leqslant N - 1$, *then*

$$\left| x - \frac{p_n}{q_n} \right| \leqslant \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}, \tag{15}$$

and, if $n > 1$, $0 < q \leqslant q_n$, *and* $p/q \neq p_n/q_n$, *then*

$$\left| x - \frac{p_n}{q_n} \right| < \left| x - \frac{p}{q} \right|, \tag{16}$$

which shows that p_n/q_n *is, among all fractions whose denominator does not exceed* q_n, *the fraction which provides the best approximation to* x.

PIPER'S ALGORITHM. *Piper's algorithm is based on* (13). *Using the example given in* (3)–(6) *we have*

$$x = \frac{826}{773} = \frac{[a_4, a_5, a_6]p_3 + p_2}{[a_4, a_5, a_6]q_3 + q_2} = \frac{(22/9)31 + 16}{(22/9)29 + 15} \tag{17}$$

where $[a_4, a_5, a_6] = [2, 2, 4] = 22/9$. *The value of the expression* $[a_4, a_5, a_6]$ *is* $2 + 1/(2 + 1/4)$, *one of the denominators of the continued fraction* (5). *Piper replaces* $[a_4, a_5, a_6] = 22/9$ *by fractions* c/d "*close*" *to* $22/9$, *calculates the value* X *of the new continued fraction through the relation*

$$X = \frac{(c/d)p_3 + p_2}{(c/d)q_3 + q_2},$$

factors the numerator and denominator of X *into their prime factors, and, if these factors can be arranged as in* (2), *calculates the error* $|X - x|$. *He considers only fractions* c/d *satisfying the rule* $cq_3 + dq_2 \leqslant 120 \cdot 120 = 14{,}400$. *Piper states that this procedure will find the optimal solution of* (2).

We give a few preliminary results before describing Piper's algorithm more carefully. Let x be a positive rational and $x_0, x_1, \ldots, x_N$ the convergents of x. Let I_n be the open interval with endpoints x_{n-2} and x_{n-1}, $2 \leqslant n \leqslant N$, and consider the function

$$f_n(t) = \frac{tp_{n-1} + p_{n-2}}{tq_{n-1} + q_{n-2}}, \qquad t > 0. \tag{18}$$

If we compute $f_n'(t)$ and use (12) we obtain

$$f_n'(t) = (p_{n-1}q_{n-2} - p_{n-2}q_{n-1})/(tq_{n-1} + q_{n-2})^2 = (-1)^{n-2}/(tq_{n-1} + q_{n-2})^2,$$

from which it follows that f_n is either strictly increasing or decreasing on $(0, \infty)$. Hence the range of f_n is equal to I_n. If an optimal solution x^* lies in I_n, then there is a unique rational $t^* \in (0, \infty)$ for which $f_n(t^*) = x^*$. Writing $t^* = c^*/d^*$, where c^* and d^* are relatively prime, we show that $c^*p_{n-1} + d^*p_{n-2}$ and $c^*q_{n-1} + d^*q_{n-2}$ are relatively prime. Setting

$$u^* = c^*p_{n-1} + d^*p_{n-2} \qquad v^* = c^*q_{n-1} + d^*q_{n-2}$$

and using (12) again we obtain

$$\pm c^* = u^*q_{n-2} - v^*p_{n-2} \qquad \pm d^* = u^*q_{n-1} - v^*p_{n-1}.$$

From the last two equations it follows that if an integer d divides u^* and v^*, then d divides c^* and d^*. Hence u^* and v^* are relatively prime.

The optimal $x^* = u^*/v^*$ can be written as $(n_1n_3)/(n_2n_4)$. Since u^* and v^* are relatively prime and $n_j \leqslant 120$, it follows that $u^*, v^* \leqslant 120^2 = 14{,}400$.

Let x^* denote an optimal solution. We have shown that if $x^* \in I_n$, where $2 \leqslant n \leqslant N$, then there are unique, relatively prime integers c^* and d^* for which

$$x^* = \frac{c^*p_{n-1} + d^*p_{n-2}}{c^*q_{n-1} + d^*q_{n-2}}, \tag{19}$$

where $c^*p_{n-1} + d^*p_{n-2}$ and $c^*q_{n-1} + d^*q_{n-2}$ are relatively prime and not greater than 120^2. If neither I_n contains a feasible solution nor x_{n-1} is feasible, then we seek a feasible solution in I_{n-1}. Since it follows from (8) and (9) that

$$f_n(t) = f_{n-1}(a_{n-1} + 1/t) \tag{20}$$

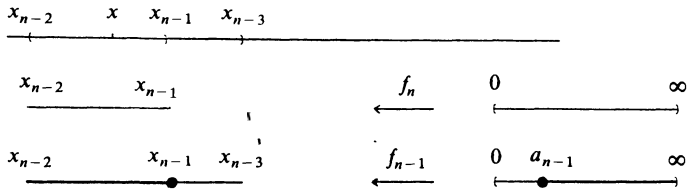
and, moreover, $x_{n-1} = f_{n-1}(a_{n-1})$, the search for c^*/d^* in the domain of f_{n-1} can be restricted to the interval $(0, a_{n-1})$.

We can now describe Piper's algorithm in detail. It is sufficient to assume that $W/w < 1$. (If $W/w = 1$, then $(20 \cdot 20)/(20 \cdot 20)$ is optimal. If $W/w > 1$, we interchange input and output.)

Piper's algorithm begins by computing the convergents of W/w. If the last convergent $x = x_N$ is feasible, then it is optimal. If x_N is not feasible, consider the function f_N and the interval I_N. We search I_N for feasible solutions by computing ratios $f_N(t) = f_N(c/d)$, where c and d are relatively prime integers, for which (21) holds. Since $p_k \leqslant q_k$ for $0 \leqslant k \leqslant N$, both constraints in (21)

$$cp_{n-1} + dp_{n-2} \leqslant 14{,}400 \qquad cq_{n-1} + dq_{n-2} \leqslant 14{,}400 \tag{21}$$

are satisfied if the second alone is satisfied. Let F_N be the set of feasible solutions in I_N together with x_{N-1}, if the latter is feasible. Next consider the function f_{N-1}. We see from (20) that the search of the domain of f_{N-1} may be restricted to the interval $(0, a_{n-1})$. Let E_{N-1} be the set of feasible solutions in I_{N-1} together with x_{N-2}, if it is feasible. From (14) it follows that if there is a feasible solution x_f in F_N at least as close to x as any solution in $E_{N-1} - F_N$, then x_f is optimal. Otherwise, let $F_{N-1} = E_{N-1} - F_N$.



This procedure is repeated until an optimal solution is found or, if $F_2 = \varnothing$, an optimal solution is chosen by finding in (x_1, ∞) the feasible ratio closest to x.

Finally, we discuss three examples given by Piper [**6**], in which the number n of teeth on a gear satisfies $20 \leqslant n \leqslant 120$, and one example from [**4**], in which $18 \leqslant n \leqslant 200$. The computations are based on a program using a computer algebra system. Piper used a table of ratios (contained in [**5**]) in his computation.

Example 1. For $W/w = 8{,}639/10{,}000$, the convergents are $0/1$, $1/1$, $(3 \cdot 2)/7$, $13/(5 \cdot 3)$, $19/(11 \cdot 2)$, $(73 \cdot 2)/(13^2)$, $(179 \cdot 5)/(37 \cdot 7 \cdot 2^2)$, $(11^2 \cdot 2^4)/(83 \cdot 3^3)$, and $(163 \cdot 53)/(5^4 \cdot 2^4)$. In this example $N = 8$ and $x_7 = (11^2 \cdot 2^4)/(83 \cdot 3^3) = (44 \cdot 44)/(83 \cdot 27)$ is optimal. The error, $|W/w - x_7|$, is 4.4×10^{-8}. The program found two other possible feasible ratios in (x_6, x_5). These correspond to $c/d = 5/2$ and $c/d = 2/3$. The ratio for $2/3$ is feasible but x_7 is closer to W/w. The high accuracy of these results is reasonable in view of Theorem C.

Example 2. For $W/w = 100{,}000/348{,}592$, the convergents are $0/1$, $1/3$, $2/7$, $(7 \cdot 5)/(61 \cdot 2)$, $107/373$, $(71 \cdot 2)/(11 \cdot 5 \cdot 3^2)$, $(509 \cdot 3)/5{,}323$, and $(5^5 \cdot 2)/21{,}787$. Here $N = 7$ and in (x_6, x_5) the program found one possible feasible ratio, the optimal solution $(63 \cdot 31)/(92 \cdot 74)$. The error is 7.4×10^{-8}.

Example 3. For $W/w = 7{,}348/10{,}000$, the convergents are $0/1$, $1/1$, $2/3$, $3/(2^2)$, $11/(5 \cdot 3)$, $(5 \cdot 5)/(17 \cdot 2)$, $(3^2 \cdot 2^2)/(7^2)$, $(19 \cdot 7)/181$, $(71 \cdot 2^3)/773$, and $(167 \cdot 11)/(5^4 \cdot 2^2)$. Here $N = 9$ and the number of feasible solutions to be factored and compared is, as Piper notes, larger than in Example 2. The program found 30 possible feasible solutions between x_8 and x_7. The optimal solution is $(79 \cdot 41)/(76 \cdot 58)$. The error is 3.6×10^{-7}.

Example 4. For $W/w = 100{,}000/517{,}220$, the convergents are $0/1$, $1/5$, $5/(13 \cdot 2)$, $(3 \cdot 2)/31$, $29/(5^2 \cdot 3 \cdot 2)$, $151/(71 \cdot 11)$, $331/(107 \cdot 2^4)$, $(241 \cdot 2)/(277 \cdot 3^2)$, $(251 \cdot 3^2)/(127 \cdot 23 \cdot 2^2)$, and $(5^4 \cdot 2^3)/(2{,}351 \cdot 11)$. Here $N = 9$ and the optimal solution is $(101 \cdot 51)/(173 \cdot 154)$ with error 1.6×10^{-8}. One other feasible solution was found in (x_8, x_7). This solution was not given in [**4**].

The author would like to thank the referee for his help.

REFERENCES

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, fourth edition, Oxford Univ. Press, 1971.
2. Heini Halberstam, An obituary for Loo-keng Hua, *The Mathematical Intelligencer*, 8 (1986) 63–65.
3. W. C. Orthwein, Determination of gear ratios, *Journal of Mechanical Design*, 104 (1982) 775–777.
4. _____, Finding gear teeth ratios with a personal computer, *Computers in Mechanical Engineering*, 4 (1985) 36–40.
5. Ray M. Page, 14000 Gear Ratios, 2nd ed., Industrial Press, NY, 1961.
6. Charles A. Piper, Approximating arbitrary gear ratios with gears, *Machinery*, 60 (1953) 162–167.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

Overdetermined Systems of Linear Equations

GARETH WILLIAMS
Department of Mathematics, Stetson University, DeLand, FL 32720

It is well known that the coefficients in a least squares fit for a given set of data points are found by "solving" an overdetermined system of linear equations. The vector of coefficients can be found by applying the pseudoinverse of the matrix of coefficients. In this paper we explore an overdetermined system of linear equations to find an appropriate "best solution." A geometrical interpretation of this solution is given. Much of the general theory of least squares fit is arrived at from a different point of view.

Consider a system of linear equations $AX = Y$, where A is an $m \times n$ matrix of rank n, with $m > n$. Such a system is said to be **overdetermined.** In general such a system will have no solution. Let us solve this system as "best we can." A is not a square matrix. There is thus no possibility of multiplying both sides of the equation $AX = Y$ by the inverse of A to get X. However we can modify the equation so that X can be isolated. Multiply both sides of the equation by A^t to get

$$A^tAX = A^tY.$$

Note that A^tA is a square $n \times n$ matrix. The rank of A^tA is known to be the same as the rank of A, namely, n [1]. Thus $(A^tA)^{-1}$ exists. Multiply both sides of this equation by $(A^tA)^{-1}$ to isolate X. We get

$$X = (A^tA)^{-1}A^tY.$$

It is instructive to ask a class at this time whether we have arrived at a contradiction. We started with a system that might not have a solution and have apparently arrived at a solution! The discussion drives home the point that extraneous solutions can arise by multiplying both sides of a matrix equation by a matrix. Denote this specific vector $\bar{X}$.

$$\bar{X} = \left[(A^tA)^{-1}A^t\right]Y.$$

The matrix $(A^tA)^{-1}A^t$ was arrived at by seeking to isolate X in the equation $AX = Y$. $(A^tA)^{-1}A^t$ is called the **pseudoinverse** [2] of A, and is denoted A^+. It is also known as the **Moore-Penrose inverse** [3] of A. $\bar{X}$ is called the **least squares solution** [2] of the system $AX = Y$.

$\bar{X}$ is not a solution to the system $AX = Y$ in the standard sense that it satisfies each equation of the system. Let us determine its significance. Consider the equation which defines $\bar{X}$, namely,

$$A^tA\bar{X} = A^tY.$$

Use the properties of matrices to rewrite this equation as follows.

$$A^tA\bar{X} - A^tY = 0$$
$$A^t(A\bar{X} - Y) = 0.$$

This equation implies that the vector $(A\bar{X} - Y)$ is orthogonal to the column space of A, which is also the range of A. $A\bar{X}$ is thus the projection of Y onto the range of A and $\bar{X}$ is the value of X such that $\|AX - Y\|$ is a minimum. The system $AX = Y$ has no solution in general; that is Y is not in the range of A. $\bar{X}$ is the best we can get; $A\bar{X}$ is the vector in the range of A that is closest to Y. The vector $(A\bar{X} - Y)$, is called the **error vector** since it gives the difference between $A\bar{X}$ and Y, and its length, $\|A\bar{X} - Y\|$, is called the **error**.

We now derive a further geometric interpretation of $\bar{X}$. Let $A\bar{X} = \bar{Y}$. We can write

$$\|\bar{Y} - Y\| = \sqrt{\Sigma(\bar{y}_i - y_i)^2}.$$

$\bar{X}$ is a solution to the system $AX = \bar{Y}$ such that $\|\bar{Y} - Y\|$, that is $\Sigma(\bar{y} - y)^2$, is a minimum.

Consider the parallel subsets corresponding to the ith equations in the two systems $AX = \bar{Y}$ and $AX = Y$. The distance d_i between two such subsets is [4]

$$d_i = \frac{|\bar{y}_i - y_i|}{\sqrt{\Sigma_j a_{ij}^2}}.$$

$\bar{X}$ is a point in this subset of $AX = \bar{Y}$, so d_i is the distance from $\bar{X}$ to the subset corresponding to the ith equation of $AX = Y$.

We now show that $\bar{X}$ is the vector, or point, which minimizes $\sqrt{\Sigma_i d_i^2}$. Normalize the system $AX = Y$ by dividing both sides of the ith equation by $\sqrt{\Sigma_j a_{ij}^2}$. The modified system is equivalent to the original system. In the new system $\sqrt{\Sigma_j a_{ij}^2} = 1$ for each i. Note that this can be accomplished by multiplying both sides of $AX = Y$ by the diagonal matrix D, where

$$d_{ii} = \frac{1}{\Sigma_j a_{ij}^2}.$$

The system is now in the equivalent form $DAX = DY$. As one would expect, and want, $\bar{X}$ is invariant under this transformation since

$$\begin{aligned}\left((DA)^t DA\right)^{-1}(DA)^t DY &= (A^t DDA)^{-1} A^t DDY \\ &= (D^2 A^t A)^{-1} D^2 A^t Y \\ &= (A^t A)^{-1}(D^2)^{-1} D^2 A^t Y \\ &= (A^t A)^{-1} A^t Y.\end{aligned}$$

In the normalized system the expression for d_i is simpler:

$$d_i = |\bar{y}_i - y_i|$$

and yields

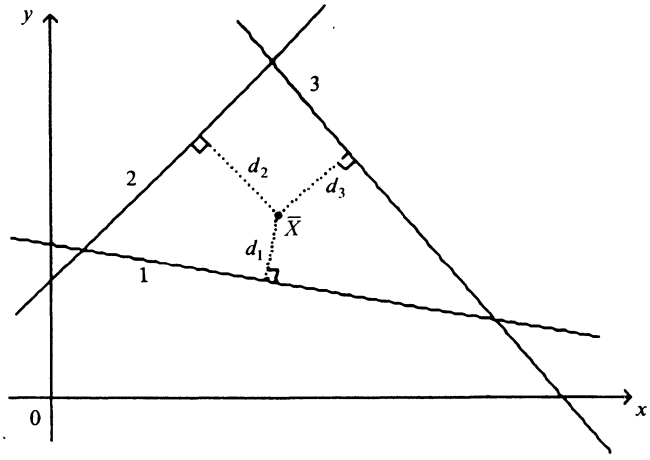
$$\|\bar{Y} - Y\| = \sqrt{\Sigma d_i^2}.$$

Since the least squares solution, $\bar{X}$, is the vector or point which minimizes $\|\bar{Y} - Y\|$ we get the following result.

The least squares solution to an overdetermined system of linear equations is the point such that the sum of the squares of the distances from the point to each of the subsets defined by the linear equations is a minimum.

Thus the least squares solution to an overdetermined system is not a solution in the sense that it necessarily satisfies every equation of the system. It is a solution in the sense that it is the value of X that comes closest to satisfying all the equations in the above geometrical way.

We illustrate this result for an overdetermined system of three equations in two variables in the figure below. The least squares solution $\overline{X}$ to the system corresponding to lines 1, 2, and 3 is the point such that $d_1^2 + d_2^2 + d_3^2$ is a minimum.



It is interesting to note the duality between the standard least squares curve discussion and this least squares point discussion. The points in the standard least squares curve discussion correspond to the subsets here and the least squares curve corresponds to the least squares point of this discussion.

REFERENCES

1. John B. Fraleigh and Raymond A. Beauregard, Linear Algebra, Addison-Wesley Publishing Company, 1987, p. 180.
2. George E. Forsythe, Michael A. Malcolm, and Cleve B. Moler, Computer Methods for Mathematical Computations, Prentice-Hall Series in Automatic Computation, 1977, p. 213.
3. R. Penrose, A generalized inverse for matrices, *Proc. Cambridge Philos. Soc.*, 51 (1955) 406–413.
4. Generalization of a result discussed in vector calculus. See, for example, Earl W. Swokowski, Calculus with Analytic Geometry, Prindle, Weber and Schmidt, 1981, p. 688.

When Quasinormal Implies Normal

DEAN HICKERSON, SHERMAN STEIN, and KENYA YAMAOKA
Department of Mathematics, University of California, Davis, CA 95616

In [7] Ore introduced the concept of a quasinormal subgroup of a group, a generalization of a normal subgroup. We will develop a condition, expressed in terms of the index, that ensures that a quasinormal subgroup is normal. The arguments suggest a variety of exercises for a course in group theory or Galois theory.

0. Definition of quasinormal subgroups. *Let H be a subgroup of a group G. If $gH = Hg$ for all $g \in G$, then H is a normal subgroup of G. It follows that for all subgroups J of G,*

$$JH = HJ. \qquad (*)$$

(For subsets $A, B \subset G$, $AB = \{ab : a \in A, b \in B\}$.) We will call a subgroup H of G that satisfies $(*)$ for all subgroups J of G a *quasinormal subgroup* of G. If H is a subgroup of G, then the following conditions are equivalent.

(0) H is quasinormal in G.
(1) $JH = HJ$ for every cyclic subgroup J of G.
(2) For every $g \in G$ and $h \in H$, there exist $r \in \mathbb{Z}$ and $h' \in H$ such that $hg = g^r h'$.
(3) For every subgroup J of G, $HJ = \langle H, J \rangle$, the subgroup spanned by H and J.

If G is finite, then these are also equivalent to

(4) For every subgroup J of G, $[\langle H, J \rangle : H] = [J : H \cap J]$.

That (0) and (1) are equivalent follows from the fact that every subgroup of G is a union of cyclic groups. Condition (2) is a translation of (1) in terms of elements; (3) depends on the fact that $HJ = JH$ if and only if HJ is a group. Condition (4) follows from the equation $|HJ| = |H||J|/|H \cap J|$, which is valid for finite subgroups H and J.

Quasinormal subgroups have been the subject of several investigations (see [**3**, **7**, **9**, **11**], and the Deskins and Venzke article in [**12**]).

1. When quasinormal implies normal. In [7] Ore proved that a quasinormal subgroup of a symmetric group is normal. In a similar spirit we show that, if H is a quasinormal subgroup of G and $[G:H]$ is a squarefree integer or twice a squarefree integer, then H is a normal subgroup of G. Incidentally, the density of such integers is $7/\pi^2 \cong 0.71$.

LEMMA 1.1. *A quasinormal subgroup H of a group G such that $[G:H]$ is prime is a normal subgroup of G.*

Proof. Assume that H is not normal in G and let $H' = a^{-1}Ha$ be a conjugate subgroup distinct from H. Let $K = HH' = H'H$. Since $[G:H]$ is prime and $H \subset K \subset G$, $K = G$. In particular, $a^{-1} = h'h$ for some $h' \in H', h \in H$. That is, $a^{-1} = a^{-1}h^*ah$ for some $h^*, h \in H$. From this it follows that $a \in H$, contradicting the assumption that $a^{-1}Ha \neq H$. Thus H is normal in G. □

For the next lemma, we use the fact that, if G is a group and $N \subset H \subset G$ are subgroups with N normal in G, then H is quasinormal in G if and only if H/N is quasinormal in G/N. The proof is left to the reader.

LEMMA 1.2. *A quasinormal subgroup H of a group G such that $[G:H] = 4$ is a normal subgroup of G.*

Proof. Assume that H is not normal in G and define H', a, and K as in the proof of Lemma 1.1. Since $H \subset K \subset G$ and $[G:H] = 4$, it follows that K is H or G or else $[K:H] = 2$. As in the proof of Lemma 1.1, the first two cases cannot occur. Thus $[K:H] = 2$, and H is normal in K. It follows that K is the normalizer of H, and consequently there are exactly two conjugates of H, namely H and H'.

Let $N = H \cap H'$. Then $N = \bigcap\{g^{-1}Hg: g \in G\}$, so N is a normal subgroup of G. Moreover, since H is quasinormal in G, $[K:H] = [H':N] = [H:N]$ and $[H:N]$ is, therefore, 2. (See FIG. 1.)

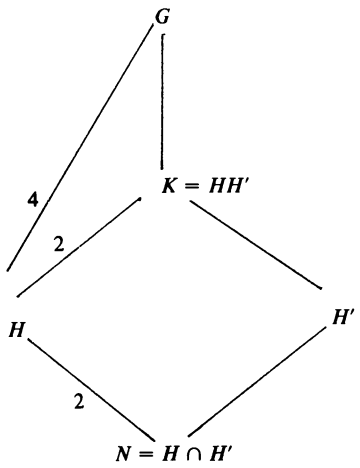


FIG. 1

We have $G \supset H \supset N$, $[G:H] = 4$, $[H:N] = 2$, and N normal in G. Thus the group G/N has order 8, H/N is quasinormal in G/N and has index 4 in G/N. Examination of the five groups of order 8 shows that a quasinormal subgroup of index 4 in a group of order 8 is normal. (Three of these groups are abelian. The other two can be checked directly or with the aid of [**10**], which displays the lattices of subgroups of small-order groups.)

Thus H/N is normal in G/N so H is normal in G, contradicting the initial assumption. From this it follows that H is normal in G.

Or, as the referee pointed out, we may argue as follows. Let $\overline{G} = G/N$ and $\overline{H} = H/N = \langle \bar{h} \rangle$. For each element $\bar{a} \in \overline{G}$ consider $\langle \bar{a}, \bar{h} \rangle$. If this group is not $\overline{G}$, then it is a group of order 2 or 4, hence abelian, and $\bar{a}$ and $\bar{h}$ commute. Thus $\overline{H}$ is normal in $\overline{G}$, hence H is normal in G. Assume, on the other hand, that $\langle \bar{a}, \bar{h} \rangle = \overline{G}$. Now, the order of $\bar{a}$ is 2, 4, or 8. If $\bar{a}$ has order 8, $\overline{G}$ is abelian. If $\bar{a}$ has order 2, condition (2) implies that $\bar{h}\bar{a} = \bar{a}\bar{h}$, so that $\overline{G}$ is abelian. If $\bar{a}$ has order 4, $\langle \bar{a} \rangle$ is a normal subgroup of $\overline{G}$. Thus $\bar{h}^{-1}\bar{a}\bar{h} = \bar{a}$ or $\bar{a}^3$. The first case implies

that $\bar{G}$ is abelian. In the second case we have $\bar{a}\bar{h} = \bar{h}\bar{a}^3$, hence $\bar{h}\bar{a}\bar{h}\bar{a} = \bar{a}^4 = 1$. Condition (2) applied to $\bar{h}$ and $\bar{h}\bar{a}$ yields $\bar{h}\bar{h}\bar{a} = \bar{h}\bar{a}\bar{h}$. Hence $\bar{h}\bar{a} = \bar{a}\bar{h} = \bar{h}\bar{a}^3$ and $\bar{a} = \bar{a}^3$, which is a contradiction. □

THEOREM 1.3. *If H is a quasinormal subgroup of a group G and $[G:H] = n$ is a squarefree integer or twice a squarefree integer, then H is a normal subgroup of G.*

Proof. We will argue by induction on n of the given type. If $n = 1$, the result is obvious. For $n = 4$ it follows from Lemma 1.2. So suppose that n is some other integer of the stated form, and assume that the claim holds for all smaller integers of the stated form. Write $n = pm$, where p is prime and does not divide m.

Suppose that H is quasinormal in G and $[G:H] = n$. Consider any element $g \in G$. Since H is quasinormal in G, $H\langle g\rangle$ is a subgroup of G and H is quasinormal in $H\langle g\rangle$. If $H\langle g\rangle \neq G$, then by the induction hypothesis, H is normal in $H\langle g\rangle$, so $Hg = gH$.

If $H\langle g\rangle = G$, then $[H\langle g\rangle:H] = n$. This implies that n is the least positive integer k such that $g^k \in H$. Let $x = g^p$ and $y = g^m$. Then the least positive integer k such that $x^k \in H$ is $n/p = m$, so $[H\langle x\rangle:H] = m$. Similarly $[H\langle y\rangle:H] = p$. Since H is quasinormal in both $H\langle x\rangle$ and $H\langle y\rangle$, the inductive hypothesis shows that $Hx = xH$ and $Hy = yH$. The fact that $(p, m) = 1$ implies that $g \in \langle x, y\rangle$; hence $Hg = gH$. □

2. When quasinormal does not imply normal. For any positive integer m that is divisible by 8 or by the square of an odd prime, we will exhibit a finite group G and a quasinormal subgroup H such that $[G:H] = m$ and H is not normal in G. The next lemma uses a group described in [**2**, p. 134; **4**, p. 187].

In its proof we will make use of two properties of commutators. Given a group G and $a, b \in G$, let $[a, b]$ denote $a^{-1}b^{-1}ab$, the commutator of a and b. Then we have ([**5**: p. 253; **1**, p. 26; **8**, p. 177]):

(i) If $[a, b]$ commutes with a, then $[a^n, b] = [a, b]^n$ for any $n \in \mathbb{Z}$.

(ii) If $[a, b]$ commutes with a and with b, then for any integer $n \geqslant 0$,

$$(ab)^n = a^n b^n [b, a]^{\binom{n}{2}}.$$

LEMMA 2.1. *For an odd prime p and $n \geqslant 3$ or for $p = 2$ and $n \geqslant 4$ let G be the group of order p^n with generators a and b and relations $a^{p^{n-1}} = 1$, $b^p = 1$, and $ba = a^{1+p^{n-2}}b$. Let $H = \langle b\rangle$. Then $[G:H] = p^{n-1}$ and H is quasinormal in G but not normal in G.*

Proof. Every element in G has a unique representation in the form $a^i b^j$ with $0 \leqslant i < p^{n-1}$, $0 \leqslant j < p$. Since $a^{-1}ba = a^{p^{n-2}}b \notin H$, H is not normal in G.

To show H is quasinormal in G, we first note the following. Since

$$ba^p b^{-1} = (bab^{-1})^p = (a^{1+p^{n-2}})^p = a^{p+p^{n-1}} = a^p,$$

we have $a^p \in Z(G)$ and $a^{p^{n-2}} \in Z(G)$. Since

$$[b, a] = b^{-1}(a^{-1}ba) = b^{-1}a^{p^{n-2}}b = a^{p^{n-2}},$$

we have $[b, a]$, $[a, b] \in Z(G)$ (since $[a, b] = [b, a]^{-1}$). (See [**1**, pp. 105, 107, and Exercise 8.2, p. 115].) So, for any $i, j \in \mathbb{Z}$, $[b^j, a^i] = [b, a]^{ij} \in Z(G)$ by (i) and

similarly $[a^i, b^j] \in Z(G)$. Also

$$[b, a]^p = (a^{p^{n-2}})^p = a^{p^{n-1}} = 1.$$

Let $g \in G$ and $h \in H$. Then $g = a^i b^j$ and $h = b^k$ for some $i, j, k \geqslant 0$. Let $r = 1 + p^{n-2}k$. By Condition (2) in Sec. 0, it suffices to show that

$$hg = g^r h.$$

By (ii),

$$g^r = (a^i b^j)^r = a^{ir} b^{jr} [b^j, a^i]^{\binom{r}{2}}.$$

Note that

$$a^{ir} = a^i (a^{p^{n-2}})^{ik} = a^i [b, a]^{ik} = a^i [b^k, a^i]$$

and that

$$b^{jr} = b^{j+p^{n-2}jk} = b^j.$$

Also the restrictions on p and n imply that p divides $\binom{r}{2}$, and

$$[b^j, a^i]^{\binom{r}{2}} = [b, a]^{ij\binom{r}{2}} = 1$$

since $[b, a]^p = 1$. Thus, $g^r = a^i [b^k, a^i] b^j$. Consequently,

$$\begin{aligned} g^r h &= g^r b^k = a^i [b^k, a^i] b^{j+k} = a^i b^k [b^k, a^i] b^j \\ &= a^i b^k (b^{-k} a^{-i} b^k a^i) b^j = b^k a^i b^j = hg. \quad \square \end{aligned}$$

LEMMA 2.2. *Let H be a quasinormal subgroup of a finite group G. If $(n, |G|) = 1$, then H is quasinormal in the group $G \times C(n)$ where $C(n)$ denotes the cyclic group of order n.*

Proof. Let $k \in G \times C(n)$ and $h \in H$. We will show that $hk = k^{r'} h'$ for some integer r' and $h' \in H$.

We have $k = ga^s$ for some $g \in G$ and integer s, where $\langle a \rangle = C(n)$. Since H is quasinormal in G, $hg = g^r h'$ for some integer r and $h' \in H$. Because $(n, |G|) = 1$, there is an integer r' such that $r' \equiv r \pmod{|G|}$ and $r' \equiv 1 \pmod{n}$. Hence $hk = hga^s = g^r h' a^s = g^{r'} h' a^s = g^{r'} a^s h' = g^{r'} a^{r's} h' = (ga^s)^{r'} h' = k^{r'} h'$, and H is quasinormal in $G \times C(n)$. $\square$

The two lemmas imply the following theorem.

THEOREM 2.3. *Let m be a positive integer that is neither a squarefree number nor twice a squarefree number. Then there is a finite group G and a quasinormal subgroup H such that $[G:H] = m$ and H is not a normal subgroup of G.*

3. Quasinormal field extensions. The notion of a quasinormal field extension can be introduced in a Galois-theory course as follows.

Let F be a field and let S be a finite-dimensional, normal extension of F. Call a field E where $F \subset E \subset S$ a *quasinormal extension* of F if for all fields K with $F \subset K \subset S$,

$$[EK: K] = [E: E \cap K]. \qquad (**)$$

(EK denotes the field spanned by E and K.) It can be shown that the definition is

independent of the normal extension S. By the "natural irrationalities theorem" (see [**6**, p. 196]), if E is a normal extension of F, it is a quasinormal extension of F.

Let $G = G(S, F)$, the Galois group of S over F. Let $H = G(S, E)$, the Galois group of S over E and let $J = G(S, K)$. Then $G(S, EK) = H \cap J$ and $G(S, E \cap K) = \langle H, J \rangle$. From $(**)$ we then have $|J|/|H \cap J| = |\langle H, J \rangle|/|H|$, for all subgroups J in G. By Condition (4) in Sec. 0, H is a quasinormal subgroup of G. Conversely, if H is a quasinormal subgroup of G, its fixed field is a quasinormal subfield of S.

In conclusion, we wish to thank the referee for calling our attention to the earlier work on quasinormal subgroups and for simplifying some of the arguments.

REFERENCES

1. M. Aschbacher, Finite Group Theory, Cambridge, N.Y., 1986.
2. W. Burnside, Theory of Groups of Finite Order, 2nd Edition, Dover, N.Y., 1911 or 1955.
3. W. E. Deskins, Quasinormal subgroups of finite groups, *Math Z.*, 82 (1963) 125–132.
4. M. Hall, The Theory of Groups, Macmillan, N.Y., 1959.
5. B. Huppert, Endliche Gruppen I, Springer, N.Y., 1967.
6. S. Lang, Algebra, Addison–Wesley, Reading, Mass., 1965.
7. O. Ore, Structures and group theory I, *Duke Math. J.*, 3 (1937) 149–174.
8. E. Schenkman, Group Theory, D. Van Nostrand, N.Y., 1965.
9. S. E. Stonehewer, Permutable subgroups of infinite groups, *Math Z.*, 125 (1972) 1–16.
10. A. D. Thomas and G. V. Wood, Group Tables, Shiva Publishing, Orpington, England, 1980.
11. J. G. Thompson, An example of core-free quasinormal subgroups of p-groups, *Math Z.*, 96 (1967) 225–227.
12. M. Weinstein, editor, Between Nilpotent and Solvable, Polygonal Pub., Passaic, N.J., 1982.

A Geometric Approach to l'Hôpital's Rule

GIANLUCA GORNI
Dipartimento di Matematica e Informatica, Università di Udine, via Zanon 6, I-33100 Udine UD, Italy

1. Introduction. Calculus textbooks usually give a few versions of L'Hôpital's rule for the evaluation of limits. The simplest case calls for hardly any proof at all: if both f and g vanish at 0 and are differentiable, with $g'(0) \neq 0$, then

$$\frac{f(x)}{g(x)} = \frac{\dfrac{f(x) - f(0)}{x - 0}}{\dfrac{g(x) - g(0)}{x - 0}} \to \frac{f'(0)}{g'(0)} \quad \text{as} \quad x \to 0.$$

The matter becomes complicated when we assume that the functions are differentiable in a neighborhood of the limit point x_0 but not necessarily in the point itself, as in the case when $x_0 = \infty$. The standard proofs then become technical and fail to provide the reader with any geometrical insight. The approach we are going to present tries to shed some light onto this side of the subject.

The idea of considering the function $f \circ g^{-1}$ isn't new, but we are not aware of any previous use of its geometrical meaning. We will also provide a hint to a parallel treatment of a "rule" for sequences.

2. A first survey. Unlike the easy instances of $f \pm g$ and $f \cdot g$, the knowledge of $\lim_{x \to x_0} f(x)$ and $\lim_{x \to x_0} g(x)$ doesn't always per se determine the existence (let alone the value) of $\lim_{x \to x_0} f(x)/g(x)$. The two nasty cases are when the two limits are either both zero or both ∞ as $x \to x_0$.

Why does this happen? In a cartesian t-u plane let us pick the point $(g(x), f(x))$ for a value of x (FIGURE 1). When $g(x) \neq 0$, the ratio $f(x)/g(x)$ is the slope of the straight line joining the origin with $(g(x), f(x))$. If we only know that $f(x)$ and $g(x)$ approach zero as $x \to x_0$, then the point $(g(x), f(x))$ gets as close as we want to the origin but total freedom is granted to the slope of the straight line. Given any fancy behavior of the ratio function $x \mapsto f(x)/g(x)$, we can easily figure out geometrically how to define f and g, both infinitesimal or both infinite, with exactly that ratio f/g.

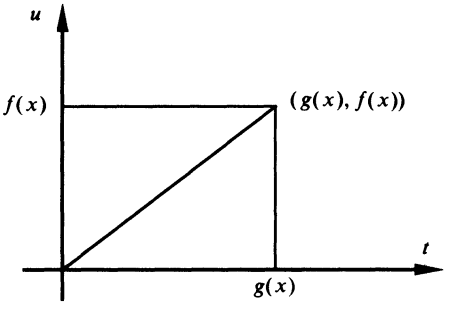


FIG. 1

So, to know what f/g does in the limit, we may need to collect more information than the mere value of the limits of f and g.

What kind of restriction on f and g might ensure that the ratio f/g tends to a limit $\lambda \in \mathbb{R}$? Let us for a moment think of x as time and consider the "trajectory" of the moving point $(g(x), f(x))$. In the case $0/0$, as a picture suggests, "$f/g \to \lambda$" seems to mean, roughly, that the straight line $u = \lambda t$ is tangent at the origin to the trajectory (FIGURE 2). The interpretation in the case ∞/∞ is less obvious.

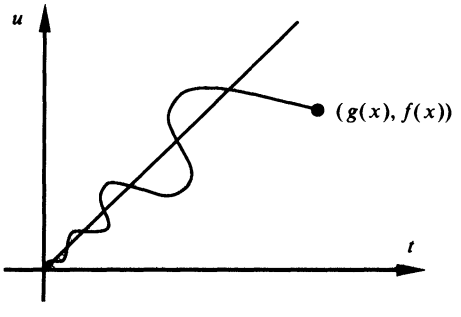


FIG. 2

Our problem is now to make precise the idea of "tangency" for a trajectory. The task is easy if the trajectory is itself the graph of a function $t \mapsto u$, because then the familiar machinery of ordinary calculus may be used. This happens, for example, whenever g is one-to-one. (We are not going to use the necessary and sufficient condition $g(x_1) = g(x_2) \Rightarrow f(x_1) = f(x_2)$.) The trajectory is then simply the graph of $h = f \circ g^{-1}$. But one-to-one is not enough. If we allow $x \mapsto g(x)$ to shuffle the order, then the origin (case $0/0$) may turn out to be surrounded by a swarm of points $(g(x), f(x))$ originated from x's well apart from x_0. In the passage from (g, f) to h, the notion of "$x \to x_0$" would be forsaken. This is not the case if g is monotone ($\lim_{t \to 0} g^{-1}(t) = x_0$ would already suffice).

Let us sum up the ideas given so far: if g is one-to-one and monotone, then the limit of f/g (in the case $0/0$) is akin to $h'(0)$, where the function h is defined as $h \equiv f \circ g^{-1}$. In the case ∞/∞ we are led to the equivalence of $\lim_{x \to x_0} f(x)/g(x)$ with $\lim_{t \to \infty} h(t)/t$.

A nitpicking detail (case $0/0$): From the very start we may have supposed that $g(x) \neq 0$ if x is close enough to x_0, just for $(g(x), f(x))$ to make sense. Then, strictly speaking, h is not defined at $t = 0$. However, when $x \to x_0$, $f(x)$ goes to zero, and so we can set $h(0) = 0 = \lim_{t \to 0} h(t) = \lim_{t \to 0} f(g^{-1}(t)) = \lim_{x \to x_0} f(x)$.

Before stating L'Hôpital's rule, we ask: What conditions can ensure the existence of $h'(t)$ at least for $t \neq 0$? Since $h = f \circ g^{-1}$, it is natural to assume f and g^{-1} differentiable; $(g^{-1})'$ exists whenever g is defined in an interval, monotone, differentiable and $g'(x) \neq 0$. If all the conditions are met, we can compute

$$h' = (f \circ g^{-1})' = (f' \circ g^{-1}) \cdot (g^{-1})' = (f' \circ g^{-1}) \cdot \left(\frac{1}{g'} \circ g^{-1} \right) = \frac{f'}{g'} \circ g^{-1},$$

$$h' \circ g = \frac{f'}{g'}, \qquad \text{i.e.,} \qquad h'(g(x)) = \frac{f'(x)}{g'(x)}. \tag{1}$$

The last formula is not at all surprising if we just think of what happens to the point $(g(x), f(x))$ when x is given the increment dx (FIGURE 3).

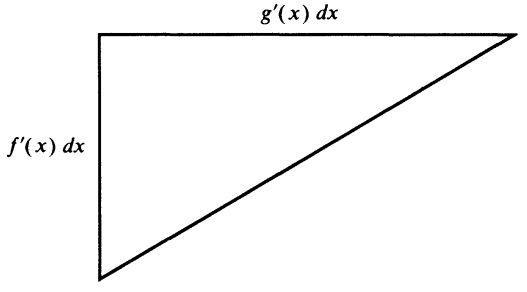


FIG. 3

3. L'Hôpital's rule. Let J be a nonempty open interval in $\mathbb{R}$, let $x_0 \in \mathbb{R} \cup \{-\infty, +\infty\}$ be one of its endpoints, and let $f, g: J \to \mathbb{R}$ be two differentiable functions, with g strictly monotone and $g'(x) \neq 0$ for all $x \in J$. Suppose that either of the next two conditions is satisfied:

(i) $f(x) \to 0$, $g(x) \to 0$ as $x \to x_0$;

(ii) $g(x) \to \pm\infty$ as $x \to x_0$ (no assumptions on f).

Then the following implication holds:

$$\lim_{x\to x_0}\frac{f'}{g'}(x) \text{ exists and equals } \lambda \in \mathbb{R}\cup\{\pm\infty\} \Rightarrow \lim_{x\to x_0}\frac{f}{g}(x) \text{ exists and equals } \lambda. \tag{2}$$

Proof. There is no restriction in assuming $g(x) > 0$, by changing the sign of both f and g if necessary. Since g is continuous, $I := g(J)$ is an interval, too, and since g is monotone, one of its endpoints is 0 in case (i) and $+\infty$ in case (ii). The function $h := f \circ g^{-1}$ is defined on I and differentiable, with derivative $h' = (f'/g') \circ g^{-1}$.

Let us first consider the easier case (i). We can extend by continuity the definition of h to $h(0) = 0$. For any $x \in J$ we can apply Lagrange's mean value theorem to h on the interval $[0, g(x)]$ (FIGURE 4),

$$\frac{h(g(x)) - h(0)}{g(x) - 0} = h'(\xi) \quad \text{with } \xi \in (0, g(x)), \tag{3}$$

that is to say,

$$\frac{f}{g}(x) = \frac{f'}{g'}(g^{-1}(\xi)). \tag{4}$$

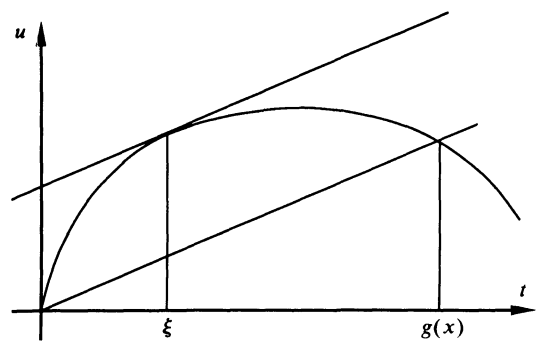


FIG. 4

Moreover, $g^{-1}(\xi)$ lies between x and x_0, because g is monotone (decreasing, in our assumption, but this doesn't matter any more). So, if we know that $f'(x)/g'(x)$ tends to λ as $x \to x_0$, we have that the common value of the two sides of (4) is as close as we wish to λ whenever x (and hence $g^{-1}(\xi)$ too) is close enough to x_0. Remark: λ in formula (2) comes out to be $h'(0)$.

Case (ii) needs more tricks, because we lack a notion of "derivative at infinity" $h'(+\infty) \equiv \lim_{t\to+\infty} h(t)/t$.

Claiming that $f(x)/g(x) \to \lambda \in \mathbb{R}$ (the case $\lambda = \pm\infty$ is easier) means that for any choice of $\lambda_1 < \lambda < \lambda_2$ we can be sure that the point $(g(x), f(x))$ lies between the two half-lines $u = \lambda_1 t$ and $u = \lambda_2 t$, $(t \geqslant 0)$ whenever x is sufficiently close to x_0.

How can we get to such a conclusion if we only know that $f'/g' \to \lambda$? Well, let us choose $\lambda_1 < \mu_1 < \lambda < \mu_2 < \lambda_2$ and $M \in J$ such that

$$(x \text{ between } M \text{ and } x_0) \Rightarrow \mu_1 \leqslant \frac{f'}{g'}(x) \leqslant \mu_2. \tag{5}$$

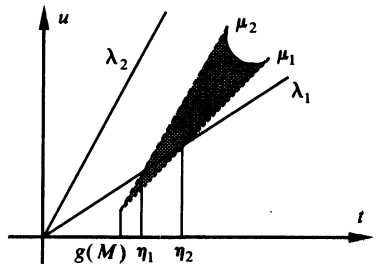


FIG. 5

If we apply the mean value theorem to h on $[g(M), g(x)]$ we see that the graph of h lies in the shaded region in FIGURE 5 for $t \geqslant g(M)$. Because of the relations between the various half-lines, we can easily check visually (and/or formally) that for $t \geqslant \bar{t} := \max\{\eta_1, \eta_2\}$, the graph of h (i.e., the point $(g(x), f(x))$ for $x \geqslant M$) falls between $u = \lambda_1 t$ and $u = \lambda_2 t$ and that is what we aimed at.

4. Extensions and counterexamples. We can somewhat relax the assumption $g'(x) \neq 0$. Suppose that g is still strictly monotone, $g'(x) = 0$ only at points isolated from each other (they may cluster against x_0), $f'(x)$ vanishes at those points too, and, finally, the function $x \mapsto f'(x)/g'(x)$ has finite limits at all these points. Then f'/g' can be extended by continuity to a function m: $J \to \mathbb{R}$, and it turns out that h is still differentiable with $h' = m \circ g^{-1}$. What happens, in terms of trajectories, is that h is differentiable on I, but as the "time" x elapses, the point $(g(x), f(x))$ may slow down, stop, and start smoothly again at certain instants. L'Hôpital's rule remains true if we replace f'/g' with m in formula (2).

We can apply such a generalization if $f'(x)$ and $g'(x)$ vanish at certain points because of a common factor that cancels out in f'/g', leaving a harmless function m, which is continuous at least at those points. But we must check that $g'(x)$ doesn't change sign (i.e., g is monotone).

Our proof breaks down miserably if we eliminate the hypothesis that g is strictly monotone, because there is no longer a reasonable h. But we may be relieved to know that formula (2) is hopeless without monotonicity. Consider FIGURE 6. We have drawn a trajectory that zigzags between two different half-lines, with horizontal cusps at the intersections (where the moving point is supposed to stop and start again smoothly in the reverse direction). The vertical span of the oscillation is

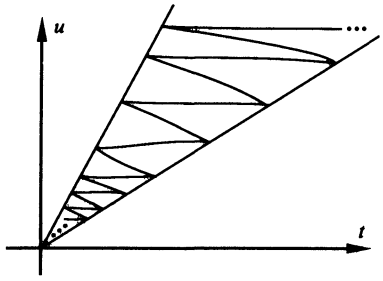


FIG. 6

supposed to shrink both when the point approaches $(0, 0)$ and infinity, in such a way that the slope of the trajectory tends to zero in both directions (things can be arranged in this fashion without much trouble). What happens analytically is that $g'(x)$ changes sign infinitely many times, $f'(x) = g'(x) = 0$ at the cusps, but the zeros cancel out in the ratio $f'(x)/g'(x)$, leaving a function that vanishes at the cusps and tends to 0 at both extremes of J. But still f/g has no limit, because it eternally oscillates between the slopes of the two straight lines. For explicit analytical examples see [1].

5. A rule for sequences. There is a theorem about the ratio of two sequences, sometimes referred to as Cesàro's rule, that closely resembles L'Hôpital's rule. Let a_n and b_n be two sequences of real numbers, b_n being strictly monotone, and suppose that either (i) they are both infinitesimal, or (ii) b_n is infinite. Then

$$\lim_{n\to\infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n} \text{ exists and equals } \lambda \in \mathbb{R} \cup \{\pm\infty\}$$
$$\Rightarrow \lim_{n\to\infty} \frac{a_n}{b_n} \text{ exists and equals } \lambda. \tag{6}$$

The proof can be constructed along the same lines as for L'Hôpital's rule. The mean value theorem will be replaced by an argument involving induction. We leave the details to the reader. See also FIGURE 7.

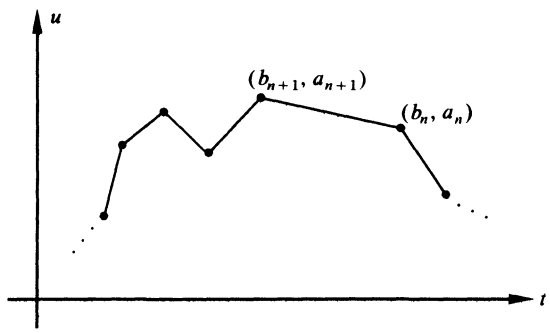


FIG. 7

Acknowledgement. This note was written while the author was visiting the Division of Applied Mathematics at Brown University, with a fellowship granted by the Italian CNR (bando 203.01.43).

REFERENCE

1. R. P. Boas, Counterexamples to L'Hôpital's rule, this MONTHLY, 93 (1986) 644–645.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, FREDERICK W. LUTTMAN, MARVIN MARCUS, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

An asterisk (*) *after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before November 30, 1990 at the MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed post card or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3391. *Proposed by David Doster, Choate Rosemary Hall, Wallingford, CT.*

Define a function on the positive integers by putting $f(1) = 1$ and setting

$$f(n + 1) = f(n) + 2 \quad \text{if} \quad f(f(n) - n + 1) = n$$

and

$$f(n + 1) = f(n) + 1 \quad \text{otherwise.}$$

(i) Find and prove a simple explicit formula for $f(n)$ involving the greatest integer function.

(ii) Show that if $f(f(n) - n + 1) \neq n$, then $f(f(n) - n + 1) = n + 1$.

E 3392. *Proposed by Antal Bege, Miercurea-Ciuc, Romania.*

Given an acute-angled triangle ABC with orthocenter H, let A_1, B_1, C_1 be the feet of the altitudes from A, B, C, respectively, and let A_2, B_2, C_2 be the feet of the perpendiculars from H onto B_1C_1, C_1A_1, A_1B_1, respectively. Prove that

$$\text{area } (\triangle ABC) \geq 16 \text{ area } (\triangle A_2B_2C)_2$$

and determine when equality holds.

E 3393. *Proposed by Bruce C. Berndt, University of Illinois at Urbana-Champaign.*

Define a function ϕ on $(-1/e, \infty)$ as follows. If $-1/e < x < \infty$, determine the unique number t in $(1/e, \infty)$ such that $x = t \log t$ and then put $\phi(x) = 1/(1 + \log t)$.

Show that $\phi^{(k)}(0) = (-k)^k$ for $k = 1, 2, 3, \ldots$.

E 3394. *Proposed by Mo Song-Qing, Institute of Applied Physics and Computational Mathematics, Beijing, People's Republic of China.*

Suppose $0 \le a_1 \le a_2 \le \cdots \le a_n$. For $2 \le k \le n-2$ consider the inequalities

$$\prod_{j=1}^{n} \frac{a_j + a_{j+1} + \cdots + a_{j+k-1}}{k} \le \prod_{j=1}^{n} \frac{a_j + a_{j+1} + \cdots + a_{j+k}}{k+1}, \tag{1}$$

and

$$\sum_{j=1}^{n} (a_j a_{j+1} \cdots a_{j+k-1})^{1/k} \ge \sum_{j=1}^{n} (a_j a_{j+1} \cdots a_{j+k})^{1/(k+1)}, \tag{2}$$

where we adopt the convention that $a_{n+j} = a_j$.

(i) Prove that (1) and (2) hold for $k = 2$.

(ii)* Prove or disprove (1) and (2) for $3 \le k \le n-2$.

E 3395. *Proposed by Hillel Gauchman, Eastern Illinois University, Charleston.*

Let A be the first orthant in n-dimensional Euclidean space E^n, i.e.,

$$A = \{x \in E^n : x = (x_1, x_2, \ldots, x_n), x_1 \ge 0, \ldots, x_n \ge 0\}.$$

Let S be a k-dimensional subspace through the origin in E^n, where $1 \le k \le n-1$, and let $S^\perp$ be the orthogonal complement to S through the origin. Prove that either S or $S^\perp$ contains a point of A other than the origin.

E 3396. *Proposed by Bruno Haible (student), Karlsruhe, West Germany.*

Suppose we define a sequence $\{x_n\}_{n=0}^{\infty}$ by putting $x_0 = 1$, $x_1 = 1/2$, and

$$(n^2+1)x_n = (2n^2 - n)x_{n-1} - (n^2 - n)x_{n-2} \quad \text{for } n = 2, 3, \ldots .$$

Prove that the sequence $\{x_n\}$ is bounded and determine $\liminf x_n$ and $\limsup x_n$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Companion Incircles of a Triangle

E 3236 [1987, 877]. *Proposed by Noam D. Elkies, Harvard University, Cambridge, MA.*

For a plane triangle $A_1A_2A_3$, call two circles within the triangle *companion incircles* if they are the incircles of two triangles formed by dividing $A_1A_2A_3$ into two smaller triangles by passing a line through one of the vertices and some point on the opposite side.

(a) Show that any chain of circles $C_1, C_2, \ldots$ such that C_n, C_{n+1} are companion incircles for every n consists of at most six distinct circles.

(b) Give a ruler and compass construction for the unique chain which has only three distinct circles.

(c) For such a chain of three circles, show that the three subdividing lines are concurrent.

Solution by Jiro Fukuta, Aichi University, Aichi, Japan, expanded by the editors. Let r be the radius of the incircle of triangle $A_1A_2A_3$, s be the semi-perimeter of $A_1A_2A_3$, and r_n be the radius of the nth circle. We may choose indices so that C_1, C_2 are companion incircles generated by a line through A_3, with C_1 in the triangle containing A_3A_1 and C_2 in the triangle containing A_3A_2. Then, successively, the circle C_n is the incircle of two triangles containing the vertex $A_{n \bmod 3}$. Thus we define A_n by $A_n = A_m$ whenever $n \equiv m \pmod 3$. Let α_n be the angle at A_n, let a_n be the length of the side opposite A_n, and let h_n be the altitude from A_n. We have $\alpha_n = \alpha_m$, $a_n = a_m$, and $h_n = h_m$. The area of triangle $A_1A_2A_3$ can be given by rs or by any of $\frac{1}{2}a_nh_n$.

In Theorem 2 of [**1**], Demir shows that successive r_n's satisfy

$$\frac{1}{r_n} + \frac{1}{r_{n+1}} - \frac{r}{r_nr_{n+1}} = \frac{2}{h_{n+2}}.$$

Multiplying this by r and subtracting it from 1, we obtain

$$\left(\frac{r}{r_n} - 1\right)\left(\frac{r}{r_{n+1}} - 1\right) = 1 - \frac{2r}{h_{n+2}}.$$

Multiplying the corresponding equations for n and $n + 2$ and dividing by the equation for $n + 1$ yields

$$\left(\frac{r}{r_n} - 1\right)\left(\frac{r}{r_{n+3}} - 1\right) = \left(1 - \frac{2r}{h_{n+1}}\right)\left(1 - \frac{2r}{h_{n+2}}\right)\Big/\left(1 - \frac{2r}{h_n}\right).$$

Since $rs = a_nh_n/2$, we have $1 - 2r/h_n = (s - a_n)/s$, so this equation becomes

$$\left(\frac{r}{r_n} - 1\right)\left(\frac{r}{r_{n+3}} - 1\right) = \frac{(s - a_{n+1})(s - a_{n+2})}{s(s - a_n)} = \tan^2 \alpha_n/2, \qquad (1)$$

where the last equality is well known to students of elementary trigonometry.

Proof of (a). Replace n by $n + 3$ in (1). Since $\alpha_{n+3} = \alpha_n$, we have $(r/r_n - 1)(r/r_{n+3} - 1) = (r/r_{n+3} - 1)(r/r_{n+6} - 1)$. Because $r_{n+3} < r$, we may cancel the common factor and obtain $r_n = r_{n+6}$, which suffices.

Proof of (b). The hypothesis is equivalent to $r_{n+3} = r_n$, so that (1) reduces to $r/r_n = 1 + \tan \alpha_n/2$. We can construct the desired circles if we can construct their centers C_1, C_2, C_3, which lie on the respective bisectors of the angles at A_1, A_2, A_3. These angle bisectors are of course concurrent at I, the incenter of triangle $A_1A_2A_3$.

Using $n = 1$, for example, we proceed to construct C_1 (see FIG. 1). From I, construct the perpendiculars to the three sides, denoting the point of intersection opposite A_i by T_i. To construct C_1, we construct the bisector of right angle A_1T_3I; its intersection with A_1I is C_1.

To verify that this is correct, let D_3 be constructed on T_3A_2 so that $T_3D_3 = IT_3 = r$, and thus IT_3D_3 is an isosceles right triangle. Then drop a perpendicular from C_1 to A_1T_3 at P_3. Using the similarity of A_1IT_3 and $A_1C_1P_3$, we have $A_1I/A_1C_1 = IT_3/C_1P_3 = r/r_1$. We want $r/r_1 = 1 + \tan \alpha_1/2$; the latter is $1 + IT_3/A_1T_3 = (A_1T_3 + T_3I)/A_1T_3 = A_1D_3/A_1T_3$, and triangles $A_1C_1T_3$ and A_1ID_3 are similar because C_1T_3 and ID_3 are parallel.

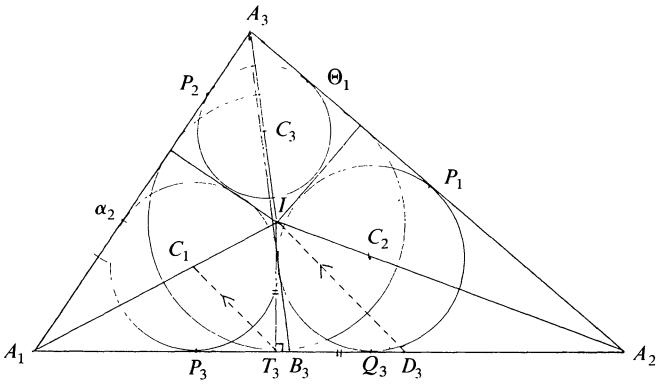


FIG. 1. Construction of unique 3 circles by ruler and compass.

Proof of (c). Let $p_i = A_{i+1}P_i$, and let the pairwise common tangent of the circles around C_i and C_{i+1} (which passes through A_{i+2}) meet the side A_iA_{i+1} at B_i. Let Q_i denote the intersection of the circle about C_{i-1} with $A_{i-1}A_{i+1}$, and let R_i denote the intersection of this circle with A_iB_i.

The semiperimeter of triangle $A_3A_1B_3$ is $s - p_1$, because $A_3B_3 = A_3R_3 + R_3B_3 = A_3P_1 + B_3Q_3$, and thus the perimeter of $A_1A_2A_3$ equals that of $A_3A_1B_3$ plus two segments of length p_1, namely A_2P_1 and A_2Q_3. By taking the ratio of the areas of $A_3A_1B_3$ and $A_3B_3A_2$ (and similarly for the rotations of the indices), we obtain

$$\frac{A_1B_3}{B_3A_2} = \frac{r_1(s-p_1)}{r_2(s-p_3)}, \qquad \frac{A_2B_1}{B_1A_3} = \frac{r_2(s-p_2)}{r_3(s-p_1)}, \qquad \frac{A_3B_2}{B_2A_1} = \frac{r_3(s-p_3)}{r_1(s-p_2)}.$$

Multiplying the three equations yields $A_1B_3 \cdot A_2B_1 \cdot A_3B_2 = B_3A_2 \cdot B_1A_3 \cdot B_2A_1$, whence Ceva's Theorem implies that A_1B_1, A_2B_2, and A_3B_3 are concurrent—at a "notable" point for a triangle, namely, the center of the inner soddy circle.

Editorial comment. Readers may note the related but significantly different result known as Malfatti's problem, which is discussed in [2] and [3]; at least one reader confused the two problems. The common tangents constructed in (c) are *not* the radical axes, and their point of concurrence is *not* the radical center.

1. Hüseyin Demir, Incircles within, *Math. Magazine* 59 (1986) 77–83.
2. A. S. Winsor, Modern Higher Plane Geometry, Boston, 1941, pp. 206–208.
3. C. J. A. Evelyn, G. B. Money-Coutts and J. A. Tyrrell, The Seven Circles Theorem, Section 3.3, Stacey International, London, 1974.

Solved also by H. Demir (Turkey) and the proposer.

A Vector Identity

E 3239 [1987, 996]. *Proposed by M. S. Klamkin, University of Alberta, Edmonton.*

Show that if **A** is any three-dimensional vector and **B**, **C** are unit vectors, then

$$[(\mathbf{A} + \mathbf{B}) \times (\mathbf{A} + \mathbf{C})] \times (\mathbf{B} \times \mathbf{C}) \cdot (\mathbf{B} + \mathbf{C}) = 0.$$

Interpret the result as a property of spherical triangles.

Solution by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.

Starting from the formula

$$(\mathbf{D} \times \mathbf{E}) \times \mathbf{F} = (\mathbf{F} \cdot \mathbf{D})\mathbf{E} - (\mathbf{F} \cdot \mathbf{E})\mathbf{D},$$

we get

$$\begin{aligned}[(\mathbf{A} + \mathbf{B}) \times (\mathbf{A} + \mathbf{C})] \times [\mathbf{B} \times \mathbf{C}] &= [(\mathbf{B} \times \mathbf{C}) \cdot (\mathbf{A} + \mathbf{B})][\mathbf{A} + \mathbf{C}] - [(\mathbf{B} \times \mathbf{C}) \cdot (\mathbf{A} + \mathbf{C})][\mathbf{A} + \mathbf{B}] \\ &= [(\mathbf{B} \times \mathbf{C}) \cdot \mathbf{A}][\mathbf{A} + \mathbf{C}] - [(\mathbf{B} \times \mathbf{C}) \cdot \mathbf{A}][\mathbf{A} + \mathbf{B}] \\ &= [(\mathbf{B} \times \mathbf{C}) \cdot \mathbf{A}][\mathbf{C} - \mathbf{B}],\end{aligned}$$

whence the result follows, since

$$(\mathbf{C} - \mathbf{B}) \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{C} \cdot \mathbf{C} - \mathbf{B} \cdot \mathbf{B} = 1 - 1 = 0.$$

Editorial comment. No two solvers gave the same "interpretation of the result as a property of spherical triangles." The proposer showed that the result is equivalent to the assertion that a great circle which bisects two sides of a spherical triangle intersects the third side 90° from its midpoint. His argument follows.

First, if we have two great circles on a unit sphere determined by the two pairs of points B', C' and B, C, then the two antipodal points of intersection of these circles lie on the line l of intersection of the planes of B', C', O and B, C, O, where O is the center of the sphere. Then a vector $\mathbf{V}$ on l is given by $\mathbf{V} = (\mathbf{B}' \times \mathbf{C}') \times (\mathbf{B} \times \mathbf{C})$. (Here $\mathbf{B}$ denotes the vector from O to the point B, etc.)

Second, let the respective midpoints of the sides of the spherical triangle ABC be A', B', C'. Then

$$\mathbf{A}' = (\mathbf{B} + \mathbf{C})/|\mathbf{B} + \mathbf{C}|, \quad \mathbf{B}' = (\mathbf{A} + \mathbf{C})/|\mathbf{A} + \mathbf{C}|, \quad \mathbf{C}' = (\mathbf{A} + \mathbf{B})/|\mathbf{A} + \mathbf{B}|$$

and $\mathbf{V} \cdot \mathbf{A}' = 0$. The equivalence now follows.

Solved also by A. Brown (Australia), J. Ferrer (Spain), R.-A. Galella, R. H. Garstang, G. N. Lewis, J.-M. Monier (France), R. A. Simon (Chile), J. H. Steelman, and the proposer.

Turning a Simple Ring into a Division Ring

E 3285 [1988, 763]. *Proposed by W. K. Nicholson, University of Calgary.*

Let R be a simple ring with unity element, i.e., a ring with unity element having no two-sided ideals other than itself and $\{0\}$. Let G be a multiplicative subgroup of the set of invertible elements of R.

(a) Under the assumption that either $x \in G$ or $1 - x \in G$ for every $x \in R$, show that R is a division ring.

(b) If, in addition to the assumption of (a), we assume that $-1 \in G$, show that $G = R - \{0\}$.

Solution by Virginia Commonwealth University Algebra Students, Richmond, VA.

Presumably it is intended that we may assume $1 \neq 0$.

An element k of R is said to be invertible if there exists an element h of R such that $hk = 1$.

LEMMA. *Under the assumptions of* (a), *if* h *and* k *are elements of* R *such that* $hk = 1$, *then* $kh = 1$.

Proof. Let $e = kh$. Then $e^2 = (kh)(kh) = k(hk)h = kh = e$, so that $(1 - e)e = 0$ and $e(1 - e) = 0$. By assumption either $e \in G$ or $1 - e \in G$. If $1 - e \in G$, the equation $(1 - e)e = 0$ and the invertibility of $1 - e$ give $e = 0$. But $e = 0$ implies $1 = hk = hkhk = hek = 0$, which is impossible. Thus we must have $e \in G$. But then the equality $e(1 - e) = 0$ and the invertibility of e give $e = 1$. Thus the lemma is proved (so that one-sided invertibility implies two-sided invertibility).

(a) Let F denote the set of noninvertible elements of R; we show that F is an ideal. Since it must be a proper ideal, the hypothesis of simplicity implies that $F = \{0\}$ and all nonzero elements are invertible (which makes R a division ring).

We first show that F is an additive group. Clearly F contains 0 and is disjoint from G. Since the negative of an invertible element is invertible, F is closed under negation. It remains to show additive closure. If $a, b \in F$ but $a + b \notin F$, then there exists c such that $c(a + b) = 1$, or $ca = 1 - cb$. By the lemma, c is invertible. This implies $ca, cb \in F$, for if cx is invertible, then x is also invertible. Now $cb \in F$ implies $1 - cb \in G$, but then $ca = 1 - cb$ contradicts $F \cap G = \varnothing$.

Finally, we show $RF \subseteq F$. We have already seen that an invertible element times a noninvertible element is not invertible, so assume $a, b \in F$. By assumption, we have $1 - a \in G$ and hence $1 - a$ is invertible, so $b - ab = (1 - a)b \in F$. This expresses ab as the difference of two elements of F, which by the preceding paragraph implies $ab \in F$. Similarly $FR \subseteq F$.

(b) It suffices to show that G contains every invertible element. Let a be an invertible element not in G; we also have $a^{-1} \notin G$. The assumption of (a) then implies that $1 - a, 1 - a^{-1}, (1 - a^{-1})^{-1} \in G$. Since

$$a = a(1 - a^{-1})(1 - a^{-1})^{-1} = (a - 1)(1 - a^{-1})^{-1} = (-1)(1 - a)(1 - a^{-1})^{-1},$$

the assumption $-1 \in G$ now implies $a \in G$, which is a contradiction. Thus G does contain every invertible element.

Editorial comment. Several readers submitted solutions to (a) using the Jacobson radical. They noted that the assumption implies that x or $1 - x$ is invertible, for all $x \in R$. By Proposition 15.15 of *Rings and Categories of Modules*, by F. W. Anderson and K. R. Fuller, this implies that $R/J(R)$ is a division ring, where $J(R)$ denotes the Jacobson radical of R. Since R is simple, this implies that R itself is a division ring. J. L. Ansorena and S. Kao noted that it is sufficient to assume only that R is semi-simple, in the sense of Jacobson. J. McCarron submitted a proof using the Chevalley-Jacobson Density Theorem. R. J. Chapman and J. H. Steelman reduced (a) to (b) by showing that if the assumption of (a) holds for any multiplicative subgroup, then it holds for any larger subgroup, such as the group of all units. J. D. O'Neill observed that if R is the ring of rational numbers and G is the set of positive rationals, then the hypothesis of (a) holds but the additional hypothesis of (b) does not; however, he showed that if char $(R) > 0$, then the hypothesis of (a) implies the additional hypothesis of (b).

Solved also by J. L. Ansorena (student, Spain), S. Beslin, C. Brueningsen & M. Leeney, D. Callan, R. Carroll (student), R. J. Chapman (United Kingdom), F. J. Flanigan, A. A. Jagers (The Netherlands), S. Kao, H. K. Krishnapriyan, C. Lanski, O. P. Lossers (The Netherlands), R. Martin (West Germany), J. McCarron (Canada), P. Morandi, J. D. O'Neill, J. H. Steelman, G. P. Wene, and the proposer.

A Quadratic Diophantine Equation in Six Unknowns

E 3304 [1989, 55]. *Proposed by Dante Vialetto, Olgiate Olona, Italy.*

Prove that there are infinitely many solutions of

$$ax^2 + by^2 = cz^2$$

in integers a, b, c, x, y, z with

$$a < b < c < x < y < z,$$

and obtain a family of solutions with as many independent parameters as possible.

Solution by C. C. Oursler, Southern Illinois University at Edwardsville. With integer parameters r, s, t, m, n, l, w, v, we obtain a solution $a = l^2w$, $b = m^2w$, $c = n^2w$, $x = mnrv$, $y = lnsv$, $z = lmtv$, by imposing the following constraints:

(a) $r < s < t$ is a primitive Pythagorean triple, (b) $m > \max\{s/(t - s), s/(s - r)\}$, (c) n and l are chosen so that $rm/s < l < m < n < tm/s$, (d) $w > 0$, and (e) $v > nw/(mr)$.

From $r^2 + s^2 = t^2$ follows $l^2(mnr)^2 + m^2(lns)^2 = n^2(lmt)^2$, which has the required form with $a = l^2$, $b = m^2$, $c = n^2$, $x = mnr$, $y = lns$, $z = lmt$. To satisfy the required inequalities, we need $l < m < n$ and also $mnr < lns < lmt$. The last inequality is equivalent to $n < mt/s$, and we require $m < n$. Thus, once the integer m is chosen, a suitable integer n can be found if the length of the interval $[m, mt/s]$ exceeds 1, that is, if $m > s/(t - s)$. Similarly, after the integer m is chosen, a suitable integer l can be found with $mr/s < l < m$, if $m > s/(s - r)$.

We can incorporate a factor of w into each of a, b, c and a factor of v into each of x, y, z. The requirement $c < x$ is satisfied if $v > nw/(mr)$.

The above family of solutions has five positive integral parameters, which appear to be independent. Thus, using the familiar parameterization of Pythagorean Triples, we can express r, s, t in terms of two positive integral parameters p and q, where p is odd and $(q, p) = 1$, by putting $r = \min\{(p + q)^2 - q^2, 2(p + q)q\}$, $s = \max\{(p + q)^2 - q^2, 2(p + q)q\}$, $t = (p + q)^2 + q^2$. The only restriction on the parameter m is a lower bound and thus m can be expressed in terms of a positive integral parameter. The parameter w is an arbitrary positive integer, while the only restriction on the parameter v is a lower bound and so v can be expressed in terms of a positive integral parameter.

Editorial comment. A referee provided the one-parameter family of solutions $a = u^2$, $b = (u + 1)^2$, $c = (u + 2)^2$, $x = 3(u + 1)(u + 2)$, $y = 4u(u + 2)$, $z = 5u(u + 1)$, where u is an integer greater than 3. This is easily seen to be a special case of the above solution.

Solved also by Serge Gaignoux.

ADVANCED PROBLEMS

6634. *Proposed by R. W. Zeamer, University College, London, England.*

Let S be the set of real numbers of the form

$$\sum_{n=2}^{\infty} \varepsilon_n/n!,$$

where $\varepsilon_n = 0$ or 1. Let K be the field generated by S. Is it true that $K = \mathbb{R}$?

6635. *Proposed by J. Michael Steele, Princeton University, Princeton, NJ.*

Suppose $\theta_1, \theta_2, \theta_3, \ldots$ are independent random variables each uniformly distributed on $[0, 2\pi]$. Let $\theta_{n:1}, \theta_{n:2}, \ldots, \theta_{n:n}$ denote the order statistics for $\{\theta_i : 1 \leqslant i \leqslant n\}$, i.e., $\theta_{n:1}, \theta_{n:2}, \ldots, \theta_{n:n}$ are the numbers $\theta_1, \theta_2, \ldots, \theta_n$ arranged so that $\theta_{n:1} \leqslant \theta_{n:2} \leqslant \cdots \leqslant \theta_{n:n}$. Adopting the convention that $\theta_{n:n+1} = \theta_{n:1}$, we put

$$\begin{aligned} S_n &= \sum_{i=1}^{n} \{\cos \theta_{n:i+1} + \cos \theta_{n:i}\}\{\sin \theta_{n:i+1} - \sin \theta_{n:i}\} \\ &= \sum_{i=1}^{n} \sin(\theta_{n:i+1} - \theta_{n:i}), \end{aligned}$$

so that S_n is twice the area of the polygon with vertices

$$(\cos \theta_{n:i}, \sin \theta_{n:i}), \qquad 1 \leqslant i \leqslant n.$$

Show that

(i) $0 \leqslant S_1 \leqslant S_2 \leqslant S_3 \leqslant S_4 \leqslant \cdots \leqslant 2\pi$, and

(ii) $S_n \to 2\pi$ as $n \to \infty$, with probability one.

6636. *Proposed by F. Schmidt, Bryn Mawr College, Bryn Mawr, PA.*

If G is a finite group, let $f(G)$ be the average order of an element of G, i.e.,

$$f(G) = \sum_{g \in G} \operatorname{ord}(g)/\operatorname{ord}(G),$$

where $\operatorname{ord}(g)$ is the order of the group element g and $\operatorname{ord}(G)$ is the number of elements in G.

(i) Show that if G is a nilpotent group of order n, then $f(G) \leqslant f(Z_n)$, where Z_n is the cyclic group of order n.

(ii)* Is it true that if G is any group of order n, then $f(G) \leqslant f(Z_n)$?

SOLUTIONS OF ADVANCED PROBLEMS

Points on the Euler Line

6557 [1987, 884]. *Proposed by Clark Kimberling, University of Evansville.*

Let Γ denote the circumcircle of a triangle ABC. Let A' be the point, other than A, where the A-median of ABC meets Γ. Let A'' be the point, other than A,

where the A-altitude of ABC meets Γ. Similarly define B', C', and B'', C''. Let DEF be the tangential triangle of ABC (e.g., D is the point where the line tangent to Γ at B meets the line tangent to Γ at C). Define (as usual) the Euler line of ABC to be the line containing the orthocenter, the circumcenter, and the centroid of ABC. Prove that the lines DA', EB', FC' and the lines DA'', EB'', FC'' concur in points that lie on the Euler line of ABC.

Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands. Let O be the center and R the radius of Γ. Let M denote the circumcenter of $\triangle DEF$ and Ω the circumcircle of this triangle. We now introduce some further notation, and some facts of advanced Euclidean geometry.

1. If U, V and W are in this order the midpoints of $\overline{BC}$, $\overline{CA}$ and $\overline{AB}$ we have
$$\overline{OU} \times \overline{OD} = \overline{OV} \times \overline{OE} = \overline{OW} \times \overline{OF} = R^2.$$
It follows that the inversion (O, R^2) transforms Ω into the nine-point-circle ν of $\triangle ABC$. The center of ν is therefore on OM. *Hence the Euler line of $\triangle ABC$ coincides with OM.*
2. Let G and N be the points of Gergonne and Nagel, respectively, of a triangle T. A well-known theorem states that the isogonal conjugates of G and N with respect to T are the internal and the external center of similitude of the incircle and the circumcircle of T.

We are now ready to prove the result. Since the symmedians of a triangle pass through the poles of the sides with respect to the circumcircle, the lines AD, BE and CF have the point L of Lemoine in $\triangle ABC$ in common. With respect to $\triangle ABC$ these lines are isogonal conjugates of AA', BB', and CC' respectively. Hence they are isogonal conjugates of DA', EB' and FC' in $\triangle DEF$. It follows that DA', EB' and FC' are concurrent in a point S, the isogonal conjugate of L with respect to $\triangle DEF$.

Since L is the Gergonne-point of $\triangle DEF$, the point S is (by 2) the internal center of similitude of Γ and Ω and as such on OM, that is (by 1) on the Euler-line of $\triangle ABC$.

Let A_1, B_1 and C_1 be the diametral points on Γ of A, B and C respectively. Let X be the point of contact of EF with the excircle Γ_1 of $\triangle DEF$ having its center on DO. Similarly define Y and Z. Now AA'' and AA_1 are isogonal conjugates in $\triangle ABC$. It follows that DA'' and DA_1 are isogonal conjugates in $\triangle DEF$. Also, D is the external center of similitude of Γ and Γ_1. Thus D, A_1 and X are collinear. Hence DA'' and DX are isogonal conjugates with respect to $\triangle DEF$. This is also true of EB'' and EY as well as FC'' and FZ. The lines DX, EY and FZ are concurrent, and their common point N is Nagel's point in $\triangle DEF$. We conclude (by 2) that DA'', EB'' and FC'' meet at the external center of similitude T of Γ and Ω and thus on the Euler line of $\triangle ABC$.

Editorial Comment. L. Kuipers (Switzerland) and the proposer both provided somewhat less erudite solutions based on the mechanism of "homogeneous trilinear coordinates." For example, the centroid has coordinates (a^{-1}, b^{-1}, c^{-1}) where a, b, and c designate the sides of ABC. It turns out that DA', EB', and FC' concur at
$$\left(a\{-a^4 + b^4 + c^4\}, b\{a^4 - b^4 + c^4\}, c\{a^4 + b^4 - c^4\}\right)$$

while DA'', EB'', FC'' concur at

$$(a\{-a' + b' + c'\}, b\{a' - b' + c'\}, c\{a' + b' - c'\})$$
$$\sim \left(\frac{\cos 2A}{\cos A}, \frac{\cos 2B}{\cos B}, \frac{\cos 2C}{\cos C}\right)$$

where $a' = \sin^2 2A$, etc.

Solved also by M. Orlowski and M. Pachter (South Africa), Martine Pages (France), Admiral Jose Yusty Pita (Spain), Raul A. Simon (Chile), an anonymous contributor, and the proposer.

The Rediscovery of Susan N. Brown's Integral

6572* [1988, 461]. *Proposed by David L. Book, Naval Research Laboratory, Washington, DC.*

Show that

$$\int_0^\infty \{1 - e^{-q(t)}\} \frac{dt}{t^{3/2}} = \pi,$$

where

$$q(t) = \frac{1}{\pi} \int_0^\infty \ln(1 + st) \frac{ds}{1 + s^2}.$$

This result appeared in an article in *Zhurnal Eksperimental'noi i Teoretecheskoi Fiziki*, 90 (1986) 330 by S. E. Esipov and I. B. Levinson, who stated that they had obtained it numerically and verified it to fourteen significant figures. (At the request of the editors, the Symbolic Computation Group of the University of Waterloo has verified the result to over forty significant figures.)

Solution I by Joseph Oesterlé, Université de Paris VI, France. The function q has a holomorphic continuation to the domain $U = \mathbb{C} - \mathbb{R}^-$; it is defined by the integral formula

$$q(t) = \frac{1}{\pi} \int_0^\infty \ln(1 + st) \frac{ds}{1 + s^2} \quad (t \in U) \tag{1}$$

where $\ln(re^{i\theta}) = \ln r + i\theta$ if $-\pi < \theta < \pi$. By differentiation under the integral sign one shows that the derivative of q is given by

$$q'(t) = \frac{1}{1 + t^2}\left(\frac{t}{2} - \frac{\ln t}{\pi}\right) \quad (t \in U). \tag{2}$$

The function q satisfies the functional equation

$$q(t) = q\left(\frac{1}{t}\right) + \frac{1}{2} \ln t \quad (t \in U). \tag{3}$$

This follows from the fact that both sides of (3) have the same derivative on U and the same value at $t = 1$.

It is clear from (1) that $\lim_{t \to 0, t > 0} q(t) = 0$. Combining this result with (2), one deduces that

$$|q(t)| \leq -|t| \ln |t| \quad \text{for } t \in U, |t| \text{ small enough.} \tag{4}$$

From (3) and (4) it follows that

$$e^{q(t)} \sim \sqrt{t} \quad (t \in U, |t| \to \infty) \tag{5}$$

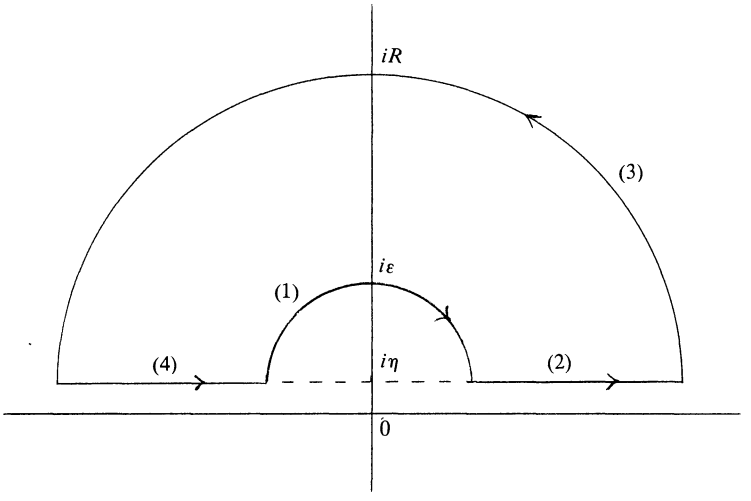
where $\sqrt{re^{i\theta}} = \sqrt{r}e^{i\theta/2}$ if $-\pi < \theta < \pi$.

The function q satisfies a second functional equation:

$$q(t) + q(-t) = \ln(1 - it) \quad (\operatorname{Im} t > 0), \tag{6}$$

valid only for t with *positive* imaginary part. This follows from the fact that both sides of (6) have the same derivative on the upper half plane, and the same limit at $t = 0$.

To solve the problem, we shall integrate $t^{-3/2}\{1 - e^{q(t)}\}\,dt$ on the contour represented in the figure (which depends on parameters η, ε, R), then take the



imaginary part of the result and successively let $\eta \to 0$, $\varepsilon \to 0$, $R \to \infty$. Let us denote by A_i, for $1 \le i \le 4$, the limit of the contribution made by the part of the contour denoted by (i) in the figure. These values are given by:

$A_1 = 0$, by formula (4);

$A_2 = 0$, because we have taken the imaginary part of the integral;

$A_3 = -\pi$, by formula (5);

$$A_4 = \int_{-\infty}^{0} \operatorname{Im}\left(\frac{1 - e^{q(t+i0^+)}}{(t + i0^+)^{3/2}}\right) dt = \int_0^{\infty} \operatorname{Im}\left(\frac{1 - (1 + it)e^{-q(t)}}{-it^{3/2}}\right) dt$$

$$= \int_0^{\infty} \frac{1 - e^{-q(t)}}{t^{3/2}}\, dt$$

by formula (6). Cauchy's theorem implies $A_1 + A_2 + A_3 + A_4 = 0$, which yields the result.

Solution II by J. Boersma, Eindhoven University of Technology, The Netherlands. From the derivative

$$q'(t) = \frac{1}{\pi}\int_0^\infty \frac{s\,ds}{(1+st)(1+s^2)} = -\frac{\log t}{\pi(1+t^2)} + \frac{t}{2(1+t^2)},$$

we deduce for $q(t)$ the alternative representation

$$q(t) = -\frac{1}{\pi}\int_0^t \frac{\log s}{1+s^2}\,ds + \frac{1}{4}\log(1+t^2).$$

By introducing the notation

$$m(t) = -\frac{1}{\pi}\int_0^t \frac{\log s}{1+s^2}\,ds,$$

the integral under consideration can be rewritten as

$$I = \int_0^\infty \left[1 - \frac{e^{-m(t)}}{(1+t^2)^{1/4}}\right]\frac{dt}{t^{3/2}}.$$

The evaluation of the latter integral is contained in the solution of the integral equation

$$f(x) = (2\pi)^{-1}\int_0^\infty (\log|x-t|)f(t)\,dt + x^{-1/2}, \qquad (0 < x < \infty)$$

as presented by Brown [**2**] and Boersma [**3**]. Indeed, by combining formulas (36a) and (3) of [**3**] and by using the integral equation with $x \to 0$, one has (in Boersma's notation)

$$I = -(2\pi)^{1/2} iF(0) = -(2\pi)^{-1/2}\int_0^\infty (\log t)f(t)\,dt$$

$$= -(2\pi)^{1/2}\lim_{x\to 0}\left[f(x) - x^{-1/2}\right].$$

The value of the latter limit is found from formulas (7.1) and (7.5a) of [**2**] to be $-(\pi/2)^{1/2}$. Thus $I = \pi$.

Editorial comments. Brown's original evaluation uses the Weiner-Hopf technique, and was done in response to a challenge by Prof. K. Stewartson. The problem arose from the study of a particular case of the Navier-Stokes equation (see [**1**]).

The solutions provided by Paul R. Chernoff, J. Koekoek (The Netherlands), William A. Newcomb, and Tomas Schonbek were similar in nature to Oesterlé's. H. K. Kuiken (The Netherlands) gave a solution quite similar to Boersma's. Kim McInturff observed that the integral is $\lim_{N\to\infty} S_N$, where

$$S_N = \sum_{k=1}^{N} \frac{(-1)^{k-1}}{k!} I_k$$

and

$$I_k = \int_0^\infty \frac{q^k(t)}{t^{3/2}}\,dt.$$

McInturff showed that $I_1 = \sqrt{2}\pi \doteq 4.4429$ and (by a *much* lengthier calculation) that

$$I_2 = 4\sqrt{2}G + \sqrt{2}\pi \ln 2 - \frac{\sqrt{2}\pi^2}{4} \doteq 4.7716$$

where G is Catalan's constant. The average of S_1 and S_2 is 3.24997..., which is slightly closer to π than is 3. This does not solve the problem, but leads us to wonder whether S_N converges to π as rapidly as the Nth partial sum of the Leibniz series, and whether it is amenable to various convergence acceleration techniques.

REFERENCES

1. A. I. Van de Vooren and A. E. P. Veldman, Incompressible viscous flow near the leading edge of a flat plate admitting slip, *J. Engrg. Math.* 9 (1975) 235–249.
2. S. N. Brown, On an integral equation of viscous flow theory, *J. Engrg. Math.*, 11 (1977) 219–226.
3. J. Boersma, Note on an integral equation of viscous flow theory, *J. Engrg. Math.*, 12 (1978) 237–243.

Functions Leaving a Distribution Unchanged

6573 [1988, 461]. *Proposed by Gérard Letac, Université Paul Sabatier, Toulouse, France.*

Suppose X is a real random variable which is normally distributed with mean 0 and variance 1. Suppose f and g are functions from $\mathbf{R}$ to $\mathbf{R}$ such that, for each b in $\mathbf{R}$, the random variable $f(X + b) - g(b)$ is also normally distributed with mean 0 and variance 1. Prove that there exists ε in $\{-1, 1\}$ and c in $\mathbf{R}$ such that $f(x) = \varepsilon x + c$ almost everywhere.

Editorial remark. The published version of this problem erroneously had $F(X + b)$ in place of $f(X + b)$.

Solution by Stephen Portnoy, University of Illinois, Urbana. It is given that

$$X \sim \text{Normal}(0, 1),$$

where " $\sim$ " means "is distributed as." If there exist real functions f and g such that

$$f(X + b) \sim X + g(b)$$

for all real b, we can conclude that $f(x) = \varepsilon x + c$. In fact, this conclusion holds whenever X has a distribution with a density f_X which is sufficiently smooth.

THEOREM. *Suppose X has a distribution with a finite variance and a density f_X which is strictly positive, continuous, and bounded. Suppose that there are real functions f and g such that for all b,*

$$f(X + b) \sim X + g(b). \tag{1}$$

Then $f(x) = \pm x + c$ almost everywhere (for some c).

Remarks. (1) If the density function f_X is not smooth, the conclusion need not hold. For example, let $X \sim \text{Uniform}(0, 1)$ and let f be periodic with period 1 and

with $f(x) = x$ for $x \in [0, 1)$. Then $f(X + b) \sim X$ for any b. In fact, this will also hold if f is periodic with period $\frac{1}{2}$ and with $f(x) = 2x$ for $x \in [0, \frac{1}{2})$.

(2) The proof below is elementary for the most part and based on a straight-forward "change of variable" calculation. There are a few measure theoretic technicalities, but the basic idea given in parts (i) and (ii) below is quite simple.

Proof. (i) First assume that f is continuously differentiable. The first step is to compute the density of $f(X + b)$ and show that f is in fact monotonic. To compute the cumulative distribution function for $f(X + b)$, let $U = \{x: f'(x) \neq 0\}$. Since $f(X)$ has a density (from (1)), U^c has measure zero. Also, U is a countable union of intervals which may be taken to be maximal: if $(a, b) \subset U$ define $a^* = \inf\{a' \leq a: (a', b) \subset U\}$ and similarly for b^*. Then $f'(a^*) = f'(b^*) = 0$ and f is monotonic on (a^*, b^*). Given real y, let x_i be such that either $f(x_i) = y$ or x_i is one of the endpoints of a maximal interval. Let f_i^{-1} be a local inverse for f on an interval containing y. Now compute the cumulative distribution function for $f(X + b)$ (that is, $P\{f(X + b) \leq y\}$) by considering the contribution from each maximal interval $(a^*, b^*) \subset U$. If $x_i = f_i^{-1}(y) \in (a^*, b^*)$, the contribution is $F_X(x_i - b) - F_X(a^* - b)$ if $f(x + b)$ is increasing on (a^*, b^*), and is $F_X(b^* - b) - F_X(x_i - b)$ if $f(x + b)$ is decreasing on (a^*, b^*). Here F_X is the distribution function of X. Otherwise, the contribution is zero if $f(x + b) > y$ on (a^*, b^*), and is $F_X(b^* - b) - F_X(a^* - b)$ if $f(x + b) < y$ on (a^*, b^*). Therefore, by (1)

$$\sum_i \varepsilon_i F_X(x_i - b) = F_X(y - g(b)) \tag{2}$$

where $x_i = f_i^{-1}(y)$, and ε_i is zero, one, or minus one as described above. Upon differentiating (2), we find that the density satisfies

$$\sum_i f_X(f_i^{-1}(y) - b)\left|\frac{d}{dy} f_i^{-1}(y)\right| = \sum_i \frac{f_X(x_i - b)}{|f'(x_i)|} = f_X(y - g(b)).$$

Now if f were not monotonic, there would be y^* which would be a local maximum for f; that is, for some i, $f'(x_i^*) = 0$. Letting $y \to y^*$ (and $x_i \to x_i^*$), the left-hand side above would tend to $+\infty$ while the right-hand side remains bounded. This contradiction shows that f must be monotonic.

(ii) It will now be shown that the conclusion of the result holds if f is monotonic. From (2), for any x and b,

$$F_X(x - b) = F_X(f(x) - g(b)) \quad \text{or} \quad F_X(z) = F_X(f(z + b) - g(b))$$

for any z and b. Differentiating with respect to b, we obtain

$$0 = f_X(f(z + b) - g(b))\{f'(z + b) - g'(b)\}.$$

Hence, $f'(z + b) = g'(b)$ for all z and b. Thus (choose $b = 0$) f' is constant and f is linear.

(iii) Finally, we show that f can be taken continuously differentiable. Since $E|X| < +\infty$, formula (1) with $b = 0$ yields

$$\int |f(x)| f_X(x)\, dx < +\infty.$$

Thus, f is integrable on any interval. Now, for any u and b, $f(X + u + b) \sim X +$

$g(u + b)$ (from (1)). Hence,

$$\int f(X + u + b)\phi(u)\,du \sim X + \int g(u + b)\phi(u)\,du, \tag{3}$$

where ϕ is any density. In particular, take ϕ to have compact support and uniformly bounded, continuous second derivative. Define

$$\tilde{f}(x) = \int f(x + u)\phi(u)\,du, \qquad \tilde{g}(x) = \int g(x + u)\phi(u)\,du. \tag{4}$$

Then, by dominated convergence,

$$\tilde{f}''(x) = \int f(v)\phi''(v - x)\,dv,$$

which is finite. Thus $\tilde{f}$ is continuously differentiable. Also, from (3), $\tilde{f}(X + b) \sim X + \tilde{g}(b)$. Therefore, from parts (i) and (ii), $\tilde{f}$ is linear. It remains to show that this implies f is also linear.

Choose $\phi(u) = \phi_t(u) = u^2(t - u)^2$ for $0 \le u \le t$ (and zero otherwise). By (4) (and dominated convergence), we have for all x and t that

$$\begin{aligned} 0 = \tilde{f}''(x) &= \int_x^{x+t} f(u)\phi''(u - x)\,du \\ &= \int_x^{x+t} f(u)\{2t^2 - 12t(u - x) + 12(u - x)^2\}\,du. \end{aligned}$$

Differentiation with respect to t yields

$$(2t^2)f(x + t) = -\int_x^{x+t} f(u)\{4t - 12(u - x)\}\,du \quad \text{a.e.} \tag{5}$$

Since f is integrable on intervals, the right-hand side of (5) is continuous. So f can be taken continuous everywhere (by perhaps redefining it on a set of measure zero). Since f is continuous, the right-hand side of (5) is continuously differentiable. Thus, differentiation of (5) with respect to x shows that f is twice continuously differentiable. Therefore, for any density ϕ with compact support and uniformly bounded, continuous second derivative, integration by parts gives

$$0 = \tilde{f}''(x) = \int f''(x + u)\phi(u)\,du.$$

As a consequence, $f''(x + u) = 0$ for all u. Hence, f is linear; that is, $f(x) = ax + c$. Finally, to show that $a = \pm 1$, note that (1) implies $\operatorname{Var} f(X + b) = \operatorname{Var} X$, or $\operatorname{Var} X = \operatorname{Var}(aX + c + b) = a^2 \operatorname{Var} X$. Thus, $a^2 = 1$ and the result follows.

Paul R. Chernoff provided a nice solution of the problem, under the original hypotheses, by complex variable techniques. The proposer gave a proof using moment generating functions and covariance matrices. Two incomplete solutions were received.

Fluid Flowing on a Sphere

6578 [1988, 666]. *Proposed by S. T. Stefanov, Sofia, Bulgaria.*

A fluid flows on a 2-sphere S with continuous velocity vector $\bar{v}(x, t)$. Let $T(x, t)$ be the temperature of the fluid at the point $x \in S$ at the time t. Assume that

$T(x,0) < T_0$ and $T(x,1) > T_0$ for all $x \in S$ and some T_0. Prove that there exist $x_0 \in S$ and $t_0 \in (0,1)$ such that $T(x_0,t_0) = T_0$ and $\bar{v}(x_0,t_0) = 0$.

Solution by John A. Frohliger, St. Norbert College, De Pere, Wisconsin. We assume, of course, that the temperature function $T(x,t)$ is continuous. Without loss of generality, S is the sphere in R^3 with center 0 and radius 1. Assume that for every t in $(0,1)$ and every x in S either $T(x,t) \neq T_0$ or $\bar{v}(x,t) \neq 0$. Identifying a point $x \in S$ with its position vector, we observe that whenever $\bar{v}(x,t) \neq 0$ the vectors x and $\bar{v}(x,t)$ are orthogonal and hence linearly independent. Define a function F such that

$$F: S \times [0,1] \to S$$

by

$$F(x,t) = \frac{(T_0 - T(x,t))x + (t - t^2)\bar{v}(x,t)}{\|(T_0 - T(x,t))x + (t - t^2)\bar{v}(x,t)\|}.$$

Then F is a homotopy from $F(x,0) = x$, the identity map on S, to $F(x,1) = -x$, the antipodal map. This is impossible, of course, since the degree of the identity map is 1, the degree of the antipodal map is -1, and homotopic maps from S to S must have the same degree. Therefore, there must exist some x_0 in S and some t_0 in $(0,1)$ such that $\bar{v}(x_0,t_0) = 0$ and $T(x_0,t_0) = T_0$.

Editorial comment. For the terminology and basic results used in the above solution, see E. H. Spanier, Algebraic Topology, McGraw-Hill, New York, 1966, pp. 175, 195–196. Peter Landweber remarks that the argument of the above solution "applies to any even-dimensional compact smooth hypersurface $M^{2n} \subset R^{2n+1}$ with nonzero Euler characteristic $\chi(M)$, since $\chi(M)$ is twice the degree of the Gauss map $M^{2n} \to S^{2n}$."

Solved also by John Bryant, W. Hensgen (Germany), Peter Landweber, John Henry Steelman, and the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Perspectives of Nonlinear Dynamics, vol. 1. By E. Atlee Jackson, Cambridge University Press, 1989, xix + 495 pp.

JOHN M. GUCKENHEIMER
Department of Mathematics, Cornell University, Ithaca, NY 14853

"Chaos" has been a trendy subject almost from the time that Li and Yorke used the word to describe complicated behavior in deterministic dynamical systems. In addition to books that try to explain the subject in non-technical terms, there have been a growing list that seek to introduce a scientific audience to the subject of nonlinear dynamics (as physicists call it) or dynamical systems theory (the mathematician's term) that provides the context within which chaotic phenomena are studied. This book is one of these. Its title gives away the background of its author. As a mathematician, I find the book frustrating. Most mathematicians will feel much more comfortable approaching the subject through ordinary differential equations texts with a modern flavor or through Devaney's book, *An Introduction to Chaotic Dynamical Systems*. Despite this attitude on my part, I think that the book deserves a closer examination. This is a time when the mathematical community seems to be examining closely what we teach to undergraduates and why. There is something about our own perspectives to be gained from taking a look at this book and comparing our scientific attitudes with those of our colleagues in physics.

The book begins with a "brief historical outline" that is a listing of scientific and mathematical achievements that the author finds notable, but the dates are imprecise and the choice is truly eclectic. When I ask myself about the purpose of this outline and the discussion of the term "nonlinear" that follows it, I come away baffled about the intended audience for this book. Clearly, it is not intended as a book for an audience lacking technical mathematical skills, and the discussion assumes too much knowledge and uses too much terminology to be accessible to students. My conclusion is that the author wrote the book for himself as a way to organize his thoughts. He invested substantial effort in learning about the subject and has given us a record of his understanding.

Technically, the book leaves a lot to be desired. It is not at all a mathematics text or systematic treatise. Subjects are discussed but not fully developed. Definitions are often incomplete, either leaving the reader to ponder the meaning of unfamiliar concepts that are used but not discussed or glossing over important points. The definitions of dimension on page 59 and structural stability on page 102 are examples. The book is maddening with its emphasis upon placing things in a historical perspective and then not getting the history quite right. The bibliography is over thirty pages, but my attempts to find the entries associated to references in the book had a poor success rate.

Despite its flaws, this book prompts difficult questions. The author states his objective in the preface as giving an introduction to the subject that is accessible to

the nonmathematician without losing much precision. Why does he feel compelled to do so? This is the question on which I would like to dwell. Mathematics strives for complete consistency and total rigor. In the process, it has created a world that is difficult for people without mathematical training to approach. Moreover, the attempts of the mathematical community to explain ideas and phenomena to a broader audience seem half-hearted and not very effective. It is left to others to write books like this one or another half dozen books with similar goals. As a result, substantial mathematical issues get either distorted or omitted in the presentation of the subject matter. Must it be this way? Are mathematicians singularly incapable of bringing mathematics to a broader public (including undergraduate students)?

By raising these questions in a rhetorical manner, it is evident that I think that we can do a better job of presenting mathematics to a wider audience. How should we begin? One step that we can take is to recognize the importance to mathematics of clear exposition that explains mathematical problems and ideas. There are historical examples (Courant and Robbins, *What is Mathematics* and Hilbert and Cohn-Vossen, *Geometry and the Imagination* are two I grew up with), but there are few recent examples of projects that have been undertaken with the intent of introducing the excitement and wonder of current mathematical discoveries to others. Gleick's book *Chaos* is written from a decidely physicist's viewpoint, yet a prominent representative of the mathematics community in Washington commented to me that Gleick's book has been one of the best things that has happened to promote mathematics to the government in decades. It seems to me that the simplest explanation as to why little effort is made to develop such exposition is that we give little recognition or reward to those who spend their time in such endeavors. In our attempts to reform calculus and the parts of the undergraduate curriculum that are not directed primarily at mathematics majors, we might also seek to motivate ourselves to bring the spirit of modern mathematics to our students in the process. In adjusting elementary courses to the background of students while standards in elementary and high school mathematics have fallen, we have failed to recognize that a single course or two will seldom overcome a decade of inadequate preparation.

Dynamical systems theory is a good example to examine in this regard. It is a subject that has always grown within an environment in which input from the physical world has played a large role. The development of many ideas has been prompted by phenomena first observed in experiments and computational studies. These phenomena can be described along with conceptual approaches to place them in a theoretical context. There is an appealing philosophical aspect to the subject in the way in which it illuminates our understanding of randomness and determinism. Discussions of these matters take us out of the realm of proving theorems, but working with these theories in a reliable fashion does require sensitivity to their mathematical structure. For example, predictions based upon the behavior of "typical" systems sometimes encounter the question as to whether behavior that occurs for values of a variable in a Cantor set of positive measure is to be regarded as typical. We have not seized the opportunity to bring this material before the public in an imaginative manner. Somehow, it seems to me that the prevailing attitudes of the mathematical community see such efforts as slightly demeaning.

What is our role as mathematicians in promoting the use of dynamical systems theory? If we insist upon narrowly circumscribed boundaries for our professional activities, then other scientists fill the gap that is left between mathematics and its use in the sciences. Their approach may clash with our mathematical values and intellectual taste. In particular, many scientists have the attitude that "truth" emerges more from successive refinements of ideas with some degree of validity than from logical reasoning. The corollary of this attitude appears to be a casual regard for accuracy and completeness in presenting mathematical concepts. One of the most compelling aspects of mathematics for me is the sense of inevitability that comes when I understand something deeply. I find satisfaction in being able to look at a phenomenon and say to myself that "it must be this way" because I can follow a chain of reasoning that starts at first principles and ends with a description of the phenomenon. I don't think that our colleagues in other sciences share these attitudes fully. This difference of values is evident to me when I approach the literature on chaos. The subject is one in which there has been a real interplay between closely reasoned mathematical theory and physical experiments, but the thrill of scientific discovery is usually given the dominant role and the underlying mathematics fades into the background. The beauty and value of the deeper understanding that comes from rigorous mathematical argument are largely forgotten.

In my opinion, mathematics would benefit if more attention were directed to the connections of mathematics with the real world and if we recognized and rewarded work that plays a significant role in making mathematical ideas available to broad audiences. We might then find books that present enlightened perspectives on modern mathematics that don't make the subject look like an impenetrable forest of jargon to outsiders.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S*, P, L**.** *The Emperor's New Mind: Concerning Computers, Minds, and The Laws of Physics.* Roger Penrose. Oxford University Pr, 1989, xiii + 466 pp, $24.95. [ISBN: 0-19-851973-7] A *tour de force* of modern science (quantum theory, computability, dynamical systems, brain physiology, quasicrystals) designed to demolish the foundation of "strong" artificial intelligence—that the brain is merely a computer—by an avalanche of evidence that consciousness is intrinsically non-algorithmic. Penrose bases much of his argument on a modern mathematical Platonism—relying in part on his own rich experience as an inventive mathematician—from which he deduces that mathematicians' judgment of mathematical truths are also non-algorithmic. Innovative, unconventional, iconoclastic, imaginative, and irresistible. LAS

General, S, P, L*. *Boojums All the Way Through: Communicating Science in a Prosaic Age.* N. David Mermin. Cambridge University Pr, 1990, xxi + 308 pp, $16.95 (P). [ISBN: 0-521-38880-5] A collection of witty and lucid essays for lay persons about physics and mathematics, mostly reprinted from *Physics Today*. "All the essays ... are informed by my desire to cut through the atmosphere of verbal dreariness with which scientists seem unnecessarily to have surrounded themselves." Essays range over quantum and relativity theories, the practice of physics, and "mathematical musings." A warm, human, and humane view of physics. (A "boojum" is a term from Lewis Carroll that the author deliberately introduced into the technical vocabulary of physics.) LAS

General, S*, P*, L.** *Does God Play Dice? The Mathematics of Chaos.* Ian Stewart. Penguin Books, 1989, 317 pp, £6.99 (P). [ISBN: 0-14-012501-9] A mathematically vigorous, popular treatment of chaos theory written with relentless energy and pun-filled wit. Using rich historical anecdotes and informative mathematical connections, Stewart traces a tale from Laplace's celestial determinism to Einstein's scepticism about quantum theory, suggesting strongly that some day we may see the unpredictability of quantum events as mere quantum chaos. LAS

General, P. *Quels Mathematiciens Pour L'an 2000?* l'Association Mathématiques A Venir (26 rue Dumeril, 75013 Paris), 1989, (P). A glossy four-color promotional brochure intended to highlight the importance of mathematics to France and the challenge "pour garantir un flux suffisant de scientifiques compétents." Based on data from 50 *lycées* gathered as a sequel to a December 1987 colloquium on this subject at l'Ecole Polytechnique. LAS

General, P*, L. *A Challenge of Numbers: People in the Mathematical Sciences.* Bernard L. Madison, Therese A. Hart. National Academy Pr, 1990, xv + 119 pp, $9.95 (P). [ISBN: 0-309-04190-2] The first report of the National Academy of Sciences Project MS 2000 containing a detailed analysis of human resources in the mathematical sciences. 45 graphs and 68 data tables document the role of mathematics in higher education and in the work place. Sample findings: the rate of growth in mathematically based occupations is about twice that for all occupations; enrollments in mathematical science courses have doubled in the last twenty years, but the increases have all been at the lower levels; attrition from the mathematical sciences is approximately half per year after the ninth grade; and the expected future supply of faculty will be insufficient to replace retirees. LAS

Elementary, S(13). *Theory and Problems of Trigonometry, Second Edition with Calculator-Based Solutions.* Frank Ayres, Robert E. Moyer. Schaum's Outline Series. McGraw-Hill, 1990, 224 pp, $9.95 (P). [ISBN: 0-07-002659-9] Revision of a 1954 review manual adding calculator use to complement logarithmic calculations. "The work with logarithms is entirely optional." Major focus is on solving triangles to calculate specific numerical values of angles, area, and lengths. LAS

Mathematics Appreciation, T(13-16), S*, L.** *Journey Through Genius: The Great Theorems of Mathematics.* William Dunham. Wiley, 1990, xiii + 300 pp, $19.95. [ISBN: 0-471-50030-5] Twelve masterpieces of mathematics from Hip-

pocrates' quadrature of the lune to Cantor's "great theorem" on non-denumerability, each set in historical context and explained in modern language ("like playing Mozart on modern instruments"). Intended for popular audiences, it offers superb source material for mathematics appreciation courses. Emphasizes biography, history, creativity, and proofs in carefully crafted vignettes—what Isaac Asimov calls "mathematics by lightning flash." LAS

Mathematics Appreciation, S* L.** *Game, Set, and Math: Enigmas and Conundrums.* Ian Stewart. Basil Blackwell, 1989, viii + 191 pp, $19.95. [ISBN: 0-631-17114-2] Twelve columns reprinted (in English) from the "Mathematical Visions" column in *Pour La Science,* the French translation of *Scientific American.* Stewart's column, written for the French edition, continues Gardner's "Mathematical Games" featured in *Scientific American.* Topics range widely from "Pascal's Fractals" to the probability of winning in tennis; the mathematical level is more demanding than Gardner's columns, but the chatty, imaginative, playful style is similar. A worthy sequel to an unsurpassed master of popular mathematical writing. Too bad it only appears in the French edition. LAS

Mathematics Appreciation, S*, L*. *Mathematical Magic Show.* Martin Gardner. MAA, 1989, 302 pp, $17.50 (P). [ISBN: 0-88385] An up-dated reprint of Gardner's 1977 collection (TR, November 1977) of *Scientific American* columns, originally published by Alfred A. Knopf. Begins with more than you ever wanted to know about "Nothing" and ends with a complementary essay on "Everything." In between are acrostics, card tricks, dragon curves, polyominoes, dice, and numerous other mathematical oddities. Includes a new introduction by Gardner fans Ronald Graham and Perci Diaconis. LAS

Mathematics Appreciation, T(13). *Mathematical Ideas, Sixth Edition.* Charles D. Miller, Vern E. Heeren, E. John Hornsby, Jr. Scott Foresman, 1990, xx + 844 pp, $29. [ISBN: 0-673-38829-8] Text for liberal arts mathematics course. In this edition a more complete introduction to calculators appears in the front of the book; introduction to the metric system is relegated to the appendix, although metric units are used in many examples and exercises; and many chapters and sections are expanded. Lots of nice topics, but few exercises go beyond mimicking the examples. (*Second Edition,* TR, March 1974; *Third Edition,* TR, June-July 1978; *Fifth Edition,* TR, October 1986, Extended Review, August-September 1988.) SB

Precalculus, T(13: 1). *Precalculus with Applications.* Stanley I. Grossman. Saunders College, 1990, xvii + 762 pp, $45.25. [ISBN: 0-03-007117-8] A well-written text whose topics include a review of basic algebra, functions and graphs, polynomials, rational functions, exponential and logarithmic functions, trigonometry, conic sections, an introduction to linear algebra, and an introduction to discrete mathematics. It has numerous examples with all steps included. Also contains many applications and brief historical readings. Each section has many exercises with answers to most odd-numbered exercises in the back. RH

Education, P*, L. *Academic Preparation In Science: Teaching for Transition From High School To College, Second Edition.* Jose P. Mestre, Jack Lochhead. Ed: Robert Orrill. College Entrance Examination Board, 1990, ix + 130 pp, (P). [ISBN: 0-87447-352-7] A major revision of the 1986 original edition, rewritten to reflect new foundations of science education provided by recent cognitive research. Focuses on the role of structured inquiry as a means of fostering higher order skills. Much of the volume relates well to mathematics—partly as a supporting skill for learning science, but primarily in terms of similarity of analyses and strategies of learning. LAS

Education, L*. *The Mathematics Curriculum: Towards the Year 2000.* Eds: John Malone, Hugh Burkhardt, Christine Keitel. Curtin U of Technology (POB U1987, GPO Perth, Western Australia 6000), 1989, 469 pp, (P). [ISBN: 0-909848-95-5] Proceedings of a Theme Group from the 1988 International Congress on Mathematical Education in Budapest, Hungary. 75 papers from over a dozen different countries arranged according to Congress sub-themes into four categories: curriculum in the year 2000, impact of computers, teachers as agents for change, and the dynamics of planned curricular change. The papers display world-wide concern for curricular renewal, but the diagnoses and prescriptions differ greatly from country to country. LAS

Education, P. *The Coordinated Implementation of National and State-by-State Reform in School Mathematics.* Association of State Supervisors of Mathematics, 1990, 212 pp, (P). Documents from a September 1989 Chicago conference of ASSM that launched state-based plans to implement the NCTM *Curriculum and Evaluation Standards for School Mathematics* (TR, June-July 1989). Contains texts of ASSM position statements, ASSM action items, and—the major part of the volume—copies of each state's action plan. LAS

Education, T(13-16: 1). *A Problem Solving Approach to Mathematics for Elementary School Teachers, Fourth Edition.* Rick Billstein, Shlomo Libeskind, Johnny W. Lott. Benjamin/Cummings, 1990, xx + 945 pp, $37.95. [ISBN: 0-8053-0390-1] Changes from the previous editions include references to the NCTM *Standards* in most chapters, a new chapter on motion geometry and tessellations, geometry activities keyed to von Hiele levels, and extensive Logo activities. Problem solving and calculator usage is expanded. Excellent text incorporating laboratory activities. (*First Edition,* TR, November 1981; Extended Review, January 1982; *Second Edition,* TR, February 1985.) MW

Education, P. *A Curriculum in Flux: Mathematics at Two-Year Colleges.* Ed: Ronald M. Davis. MAA, 1989, 48 pp, $7.50 (P). [ISBN: 0-88385-065-6] Issues and recommendations for mathematics curricula

at two-year colleges, with chapters on baccalaureate transfer programs, statistics, remediation, occupational career programs, mathematics for the computing sciences, and computing in the mathematics curriculum. A joint undertaking of AMATYC and MAA to help focus attention on the special needs of two-year colleges, where one-third of all first-year college students study mathematics. LAS

Logic, T(17: 1), L. *Logik für Informatiker.* Uwe Schöning. Reihe Informatik, Band 56. Bibliographisches Institut, 1989, 172 pp, (P). [ISBN: 3-411-14012-7] An introductory textbook on logic for computer scientists. Chapters on propositional logic, predicate logic, and logic programming. Exercises, short bibliography. DFA

Foundations, P, L*. *Pursuit of Truth.* W.V. Quine. Harvard U Pr, 1990, x + 113 pp, $17.50. [ISBN: 0-674-73950-7] A concise recapitulation of ideas on evidence, reference, meaning, intension, and truth by one of America's leading philosophers. Quine focuses on certain classes of sentences for which truth is "disquotation:" 'Snow is white' is true iff snow is white. Truth in mathematics, for Quine, inclines one towards the propositions of predicative set theory; such a view "inactivates the more gratuitous flights of higher set theory." LAS

Foundations, T(15-16: 1), S, P, L.** *Foundations and Fundamental Concepts of Mathematics, Third Edition.* Howard Eves. PWS-Kent, 1990, xvii + 344 pp, $33. [ISBN: 0-534-92163-9] A welcome reappearance of the classic *An Introduction to the Foundations and Fundamental Concepts of Mathematics* by Howard Eves and Carroll V. Newsom in a slightly revised form. JNC

Graph Theory, S(17-18), P. *Graphentheorie, Band 1: Anwendungen auf Topologie, Gruppentheorie und Verbandstheorie.* Klaus Wagner, Rainer Bodendiek. Bibliographisches Institut, 1989, xv + 252 pp, (P). [ISBN: 3-411-03226-X]

Combinatorics, P. *Sequences: Combinatorics, Compression, Security, and Transmission.* Ed: Renato M. Capocelli. Springer-Verlag, 1990, xii + 549 pp, $49.50. [ISBN: 0-387-97186-6] Forty-two papers on combinatorial algorithms, data compression, security, and transmission. Includes a Paul Erdös paper on probabilistic number theory with $2100 worth of conjectures. BC

Combinatorics, T(16-18: 1, 2), S, P, L. *Matroid Theory and its Applications in Electric Network Theory and in Statics.* András Recski. Algorithms & Combinatorics, V. 6. Springer-Verlag, 1989, xiii + 531 pp, $89.50. [ISBN: 0-387-15285-7] An introduction to matroid theory (including the necessary background about graphs and matrices), with special emphasis on its algorithmic aspects and applications in electrical engineering and statics. The breadth of the mathematical portion is between the level of Welsh's book (1977) and Lawler's work (1976). About 40% of the material is on applications (which may be skipped over if desired); most of these have never before been given in book form. Nearly 800 exercises and problems with full solutions. LCL

Number Theory, T(14-15: 1), L. *Elementare Zahlentheorie.* Friedhelm Padberg. Bibliographisches Institut, 1989, 166 pp, (P). [ISBN: 3-411-00639-0] Intended chiefly for teachers and prospective teachers of school mathematics. Exercises, solutions. JD-B

Number Theory, P, L.** *A Dictionary of Real Numbers.* Jonathan Borwein, Peter Borwein. Wadsworth, 1990, viii + 424 pp. [ISBN: 0-534-12840-8] A reverse handbook, of special functions, listing over 100,000 eight-digit decimal numbers between 0 and 1 with a crisply encoded clue as to how that particular number arises as the decimal part of the value or root of some common functions. Includes elementary and special functions, simple algebraic and transcendental constants, rational combinations of constants, common physical constants, real roots of cubic polynomials with small integer coefficients, and much more. Generated by Maple and typeset in TEX to avoid human error. An invaluable aid to mathematicians who do computer-assisted research; also makes a great coffee-table gift for those with a sense of humor. LAS

Algebra, T(15-16: 1), L.** *Galois Theory, Second Edition.* Ian Stewart. Chapman & Hall, 1989, xxx + 202 pp, $52.50; $22.50 (P). [ISBN: 0-412-34540-4; 0-412-34550-1] Changes from the *First Edition* (TR, August-September 1973; Extended Review, October 1976) include the addition of an introductory overview, a new chapter on the calculation of Galois groups, a modification of the historical section, some changes in the proofs, more examples, and some modification of the exercises. This is a very nicely written book which would serve as an excellent undergraduate text. RH

Algebra, T(15-16). *Numbers, Groups and Codes.* J.F. Humphreys, M.Y. Prest. Cambridge University Pr, 1989, xvi + 288 pp, $19.95 (P); $59.50. [ISBN: 0-521-35938-4; 0-521-35084-0] An interesting introduction to group theory and algebra written with an eye toward applications in computer science. The chapters cover number theory (congruences, factorization, public key ciphers), sets and relations (functions, finite state machines), logic (Boolean algebras, switching circuits), group theory, and elementary coding theory. Well-written with many exercises. SG

Algebra, P. *Methods of Representation Theory with Applications to Finite Groups and Orders, Volume 1.* Charles W. Curtis, Irving Reiner. Classics Library. Wiley, 1990, xxi + 819 pp, $34.95 (P). [ISBN: 0-471-52367-4] A paperback version of the 1981 hardcover edition (TR, March 1982). Still an (if not the) outstanding reference book in the field. SG

Algebra, P. *Arithmetical Completely Simple Semigroups.* E.-A. Behrens. Research & Exposition in Math., V. 17. Heldermann Verlag, 1989, vii + 135 pp, DM 48 (P). [ISBN: 3-88538-217-2] Develops a non-commutative arithmetic for partially ordered semigroups. Defines arithmetical completely simple semigroups and applies them to the integral normal

ideals in the central simple algebras over the quotient field of a Dedekind ring under a generalized multiplication. Only a basic knowledge of algebra is required as background. RH

Calculus, T(13: 2), S. *Applied Calculus, Second Edition.* S.T. Tan. PWS-Kent, 1990, xv + 775 pp, $37. [ISBN: 0-534-92135-3] Intends to "include at least one real-life application in each section (whenever feasible)." Lots of good problems; most come from business and economics, some from social and life sciences, very few from physical sciences. Changes in this edition (*First Edition,* TR, June-July 1987) include addition of a preliminary precalculus review; "self-check exercises" (with worked out solutions) at the end of each section in addition to regular exercises; additional examples and exercises (those requiring calculators are labeled); section on limits now includes discussion of one-side limits; topics of integration section has been reorganized with an earlier introduction of the definite integral and fundamental theorem; numerical integration section now includes error analysis of trapezoidal and Simpson's Rules; expanded series section includes error analysis of integral, comparison and ratio tests. As in the previous edition, there is no mention of the mean value theorem. SB

Calculus, T(13: 1). *Essential Calculus with Applications.* Richard A. Silverman. Dover, 1989, xii + 292 pp, $8.95 (P). [ISBN: 0-486-66097-4] A "corrected and slightly enlarged republication" of the original 1977 Saunders publication (TR, October 1977); now includes the section Supplementary Hints and Answers formerly issued as a separate *Instructor's Manual.* JNC

Calculus, T(13: 2-3). *Calculus.* Robert Seeley. Harcourt Brace Jovanovich, 1990, xv + 1048 pp, $44 [ISBN: 0-15-505681-6] What distinguishes this from many calculus texts is the genuine effort towards presenting motivational material *in advance of* the rigorous calculus. This comes in the form of applications set in their historical context. For example, optimization problems and motion *precede* the rigorous discussion of limits, and Taylor polynomials *begin* the chapter on sequences and series. KS

Real Analysis, T(16-17: 1). *An Introduction to the Theory of Real Functions.* Stanisław Łojasiewicz. Transl: G.H. Lawden. Wiley, 1988, ix + 230 pp, $57.95. [ISBN: 0-471-91414-2] A terse theorem-proof English version of a Polish text first published in 1964. Set in the context of metric spaces; covers Borel sets and Baire functions, Reisz's theorem, and the Radon-Nikodym theorem. Elegant, crisp, pristine; will require a well-motivated reader. LAS

Real Analysis, P, L. *Lecture Notes in Mathematics-1419: New Integrals.* Eds: P.S. Bullen, *et al.* Springer-Verlag, 1990, 200 pp, $21.80 (P). [ISBN: 0-387-52322-7] Proceedings of an August 1988 conference held at the University of Ulster in honor of Ralph Henstock, and focused on generalized integrals related to Henstock's work on the generalized Riemannian integral "That was as good as, and indeed, in many ways, far superior to Lebesgue's." LAS

Real Analysis, T(16-17: 1, 2), S. *Problems in Real Analysis: A Workbook with Solutions.* Charalambos D. Aliprantis, Owen Burkinshaw. Academic Pr, 1990, vii + 285 pp, $29.95. [ISBN: 0-12-050256-9] A collection of 456 exercises, with complete solutions, on topics in real analysis: set theory, general topology, measure theory, normed spaces, topics in Lebesgue integration, L_p spaces. Best used as a companion to the authors' book *Principles of Real Analysis* (TR, November 1982), in which all these problems appear, and *to* which many references (e.g., theorem numbers) are made. Could be a valuable study aid for readers who resist temptation (perhaps considerable—solutions appear hard on heels of problems) to peek. PZ

Real Analysis, P. *Almost Everywhere Convergence.* Eds: Gerald A. Edgar, Louis Sucheston. Academic Pr, 1989, xv + 416 pp, $49.95. [ISBN: 0-12-231050-0] The proceedings of a conference held in Columbus, Ohio, June 1988. Contains 25 papers in this expanding area of ergodic theory; some of a survey nature, others quite technical. TAV

Numerical Analysis, T(16-17: 1), L. *Numerical Analysis: A Second Course.* James M. Ortega. Classics in Appl. Math., V. 3. SIAM, 1990, xiii + 201 pp, $25.50 (P). [ISBN: 0-89871-250-5] An unabridged, corrected republication of the original 1972 edition (TR, August-September 1972; Extended Review, October 1974). Presents theoretical results pertaining to rounding error, discretization error, and convergence error. Applications of these results are given to the study of systems of linear equations and differential equations. AO

Numerical Analysis, P. *Iterative Methods for Large Linear Systems.* Eds: David R. Kincaid, Linda J. Hayes. Academic Pr, 1990, xxx + 319 pp, $49.95. [ISBN: 0-12-407475-8] Seventeen papers from a conference held October 19-21, 1988 at the University of Texas at Austin. These papers provide an overview of the state-of-the-art in the use of iterative methods for solving sparse linear systems. AO

Operator Theory, T(18: 1, 2), S, P. *Lecture Notes in Mathematics-1413: Uniqueness of the Injective III_1 Factor.* Steve Wright. Springer-Verlag, 1989, 108 pp, $13.50 (P). [ISBN: 0-387-52130-5] Based on a lecture course by the author, this volume contains an exposition of work of A. Connes (*Journal of Operator Theory,* 14 (1985)) and U. Haagerup (*Acta Math.,* 158 (1987)). Together these constitute a proof of what's promised in the title. KS

Operator Theory, P. *Matrix and Operator Extensions.* H.J. Woerdeman. CWI Tract, V. 68. Mathematisch Centrum, 1989, 158 pp, Dfl. 48 (P). [ISBN: 90-6196-379-6] Taken from the author's Ph.D. thesis, written at Vrije Universiteit in Amsterdam, it concerns the problem of extending a partially defined operator to a full operator having certain prescribed properties. LCL

Functional Analysis, T(16-17: 1, 2). *Anal-*

ysis Now. Gert K. Pedersen. Grad. Texts in Math., V. 118. Springer-Verlag, 1989, xiv + 277 pp, $49.80. [ISBN: 0-387-96788-5] Key propositions of functional analysis (topology, Banach and Hilbert spaces, spectral theory, unbounded operators, integration theory) embedded in a context to motivate the necessary abstractions that "every young analyst" should know: Norms, Operators, and Weak topologies—the "Now" of the title. "Our critics, especially those well-meaning pedagogues, should come to realize that mathematics becomes simpler only through abstraction." LAS

Functional Analysis, P. *Inverse und schlecht gestellte Probleme.* Alfred Karl Louis. BG Teubner Stuttgart, 1989, 205 pp, DM 26,80 (P). [ISBN: 3-519-02084-X]

Functional Analysis, P. *Approximation Problems in Analysis and Probability.* M.P. Heble. Math. Stud., V. 159. North-Holland (US Distr: Elsevier Science), 1989, xi + 245 pp, $92. [ISBN: 0-444-88021-6] Recent research on approximation theory with particular emphasis on probabilistic methods and applications. Introductory chapter on generalizations of the Weierstrass-Stone theorem. Appendices give brief review of theoretical background. JO

Analysis, P. *Representations of Lie Groups, Kyoto, Hiroshima, 1986.* Eds: K. Okamoto, T. Otshima. Adv. Stud. in Pure Math., V. 14. Academic Pr, 1988, 660 pp, $99.50. [ISBN: 0-12-525100-9] Proceedings of a symposium on Analysis on Homogeneous Spaces and Representations of Lie Groups in September 1986. Contains 22 papers. MLR

Analysis, T(18: 1), P. *Stochastic Calculus in Manifolds.* Michel Emery. Universitext. Springer-Verlag, 1989, x + 151 pp, $29 (P). [ISBN: 0-387-51664-6] An introduction to the Schwartz-Meyer second-order geometry and its use in stochastic calculus. Contains basics of the stochastic calculus and differential geometry, so is reasonably self-contained. In spite of this, the work is quite technical and demanding. TAV

Algebraic Geometry, P. *Lecture Notes in Mathematics-1417: Algebraic Geometry.* Eds: A.J. Sommese, A. Biancofiore, E.L. Livorni. Springer-Verlag, 1990, 320 pp, $31.20 (P). [ISBN: 0-387-52217-4] A collection of eighteen papers presented at a 1988 conference in Italy. The papers are devoted primarily to the study of hyperplane sections of projective varieties, and the role played by such sections in the classification of projective varieties by projective invariants. SG

Differential Geometry, T(16-18: 1), S, L. *Differential Forms with Applications to the Physical Sciences.* Harley Flanders. Dover, 1989, xv + 205 pp, $5.95 (P). [ISBN: 0-486-66169-5] Introduction to differential forms, aimed at weaning physicists and engineers away from tensors. A slightly revised version of the 1963 original. BC

Differential Geometry, P. *Contact Geometry and Wave Propagation.* V.I. Arnold. L'Enseignement Mathematique, 1989, 56 pp, 27 Frs. (P). Contact geometry is the odd-dimensional counterpart of symplectic geometry and is, according to the author, the natural basis for optics and the theory of wave propagation. This monograph presents, briefly, the basic results of contact geometry. JO

Differential Geometry, P. *Geometry of Manifolds.* Ed: K. Shiohama. Perspectives in Math., V. 8. Academic Pr, 1989, xv + 517 pp, $49.95. [ISBN: 0-12-640170-5] Thirty-two papers contributed by participants in the 35th Symposium on Differential Geometry held in Matsumoto, Japan in July 1988. Sections on dynamical systems, geometry of submanifolds, Lie sphere geometry, Riemannian geometry, the geometry of Laplace operators, and Yang-Mills connections. JO

Algebraic Topology, P. *Lecture Notes in Mathematics-1418: Homotopy Theory and Related Topics.* Ed: M. Mimura. Springer-Verlag, 1990, 241 pp, $21.80 (P). [ISBN: 0-387-52246-8] Proceedings of an August 1988 international conference held in honor of Hirosi Toda on the occasion of his sixtieth birthday. Opening lecture on Toda's work is by J.F. Adams; 18 other papers extend various aspects of Toda's work. LAS

Dynamical Systems, S(16-18), P, L*. *Exploring Complexity, An Introduction.* Grégoire Nicolis, Ilya Prigogine. WH Freeman, 1989, xi + 313 pp, $24.95. [ISBN: 0-7167-1859-6] One goal for this book is to provide a short introduction to methods devised over recent decades to explore complexity. From a descriptive beginning on the vocabulary of complexity (e.g., nonequilibrium, stability, bifurcation and symmetry breaking, long-range order, chaos, attractors, and fractals), the authors continue with more detailed descriptions (e.g., nonequilibrium physics and modern dynamical systems). One of the most important goals is to show that concepts of complexity are quite universal, and break through interdisciplinary barriers. The purpose is nothing less than to promote a whole new world view in which evolution and plurality (rather than stability and periodicity) are the key words. The authors (one a 1977 Nobel laureate for chemistry, the other a distinguished theoretical physicist and chemist) take every occasion to stress the importance of these concepts for better understanding the global environment in which we live. LCL

Dynamical Systems, T(18: 2). *Infinite-Dimensional Dynamical Systems in Mechanics and Physics.* Roger Temam. Appl. Math. Sci., V. 68. Springer-Verlag, 1988, xvi + 500 pp, $64. [ISBN: 0-387-96638-2] Typical of most books in this series, this volume is a reasonably self-contained, readable, graduate-level text on the subject. Includes applications of the ideas of invariant sets and attractors to a large number of physically interesting systems. First chapter has a "User's Guide" for readers interested only in learning enough for physical applications. JO

Probability, T*(14: 1). *Introduction to Probability and Its Applications.* Richard L. Scheaffer. Duxbury Adv. Ser. in Stat. & Decision Sci. PWS-Kent, 1990, x + 363 pp. [ISBN: 0-534-91970-7] A well-written

text for a one-semester, post-calculus course. Includes discrete and continuous multivariables, and examples of stochastic processes. Extensive, well-chosen exercises, clear graphics, and careful organization make this an attractive choice. TAV

Stochastic Processes. *Lecture Notes in Mathematics-1412: Stability Problems for Stochastic Models.* Eds: V.V. Kalashnikov, V.M. Zolotarev. Springer-Verlag, 1989, x + 380 pp, $33 (P). [ISBN: 0-387-51948-3] The proceedings of an international conference in Sukhumi (USSR) in the fall of 1987. The scope of the 26 papers is quite broad, as is the whole field of stability. TAV

Stochastic Processes, P. *Queueing Networks—Exact Computational Algorithms: A Unified Theory Based on Decomposition and Aggregation.* Adrian E. Conway, Nicolas D. Georganas. Computer Systems Ser. MIT Pr, 1989, xiv + 234 pp, $25. [ISBN: 0-262-03145-0] A state-of-the-art view of exact computational algorithms to analyze product-form queueing networks in equilibrium. Uses the techniques of decomposition and aggregation. Attempts to lay the ground work for future development of exact algorithms. TAV

Stochastic Processes, T?(16: 1), S(15-16), L. *Spontaneous Phenomena: A Mathematical Analysis.* Flemming Topsøe. Transl: John Stillwell. Academic Pr, 1990, xi + 182 pp, $29.95. [ISBN: 0-12-695175-6] Translation of a 1983 Danish book. Introduces the stochastic modeling of spontaneous phenomena by a detailed development, including the background physics, of a Poisson process to model radioactive decay. Includes a section of 46 problems of widely varying difficulty which can serve as the basis for a course, and a collection of 18 examples from a variety of disciplines "for further investigation." RSK

Statistics, T(14-16: 2). *Probability and Statistics for Modern Engineering, Second Edition.* Lawrence L. Lapin. Duxbury Adv. Ser. in Stat. & Decision Sci. PWS-Kent, 1990, xxii + 810 pp. [ISBN: 0-534-91654-6] Beginning (not elementary) level statistics text designed specifically for engineering students. Exercises and examples refer to engineering applications. Text assumes students have strong quantitative background, and calculus is used. A notch above most elementary statistics texts in terms of theory covered, but below most theoretical probability and statistics texts for undergraduate mathematics majors. Many theoretical results are presented (e.g., joint probability distributions, moment generating functions, gamma distribution, etc.), but most proofs are left out. Covers descriptive statistics, basic probability, random variables and distributions, sampling and sampling distributions, estimation, hypothesis testing, regression, multiple regression, analysis of variance, nonparametrics, and a nice section on special topics (control charts, acceptance sampling, reliability and life testing, Weibull distributions). Good text with new approach for engineers. JJ

Statistics, S(15-18), L. *Analyzing Quantitative Behavioral Observation Data.* Hoi K. Suen, Donald Ary. Lawrence Erlbaum Associates, 1989, xi + 359 pp, $34.50. [ISBN: 0-89859-945-8] A conceptual introduction to sampling, reliability, and validity issues in the context of analyzing human and animal behavior. Also presents ARIMA time series and Markov chain models. RWJ

Statistics, S(17), P. *Robust Regression: Analysis and Applications.* Eds: Kenneth D. Lawrence, Jeffrey L. Arthur. Statistics: Textbooks & Monographs, V. 108. Marcel Dekker, 1990, xiii + 287 pp, $89.75. [ISBN: 0-8247-8129-5] Fourteen papers on robust regression including five on robust time series analysis and two on ridge regression. Includes examples with real data. TH

Statistics, P. *Nonlinear Regression, Functional Relations and Robust Methods: Volume II, Statistical Methods of Model Building.* Eds: Helga Bunke, Olaf Bunke. Prob. & Math. Stat. Wiley, 1989, 449 pp, $69.96. [ISBN: 0-471-91239-5] Chapters on nonlinear regression models, robust linear regression, and errors-in-variables regression. Little combination of the three topics. Encyclopedic character (many topics presented quickly). Translation from German. TH

Statistics, S, P, L.** *The History of Statistics: The Measurement of Uncertainty before 1900.* Stephen M. Stigler. Harvard University Pr, 1986, xvi + 410 pp, (P). [ISBN: 0-674-40341-X] Paperback reprint of 1986 original edition (TR, August 1987). A readable, fact-filled survey focused on variability in astronomy and geodesics, probability in social science, and studies of heredity. LAS

Programming, T(14: 1, 2), S. *Accessing C: Tips from the Experts, Second Edition.* Ed: Strawberry Software. Van Nostrand Reinhold, 1989, xiii + 306 pp, $29.95 (P). [ISBN: 0-442-20662-3] Text designed to bridge the gap "between the syntax of the language and the knowledge to complete real-world applications in C." Surprisingly even quality despite a large number of contributors. Exposition is to the point and includes many helpful code examples. JO

Computer Systems, S(13-14), P, L. *Maple for the Calculus Student: A Tutorial.* Wade Ellis, Jr., Ed Lodi. Brooks/Cole, 1989, xii + 67 pp, $10.25 (P). [ISBN: 0-534-11874-7] A clear, leisurely, and friendly tutorial introduction to the rudiments of using the algebra system *Maple* for graphical, numerical, and symbolic computing. Calculus-level operations (curve plotting, symbolic differentiation, and antidifferentiation, solving differential equations, etc.) are emphasized, but not only calculus students will benefit. The booklet can serve anyone as a first quick overview of the syntax, capabilities, and "feel" of a modern computer algebra system, and as an enticement to look further. PZ

Computer Systems, P, L. LaTeX *for Scientists and Engineers.* David J. Buerger. McGraw-Hill, 1990, xvi + 198 pp, $29.95 (P). [ISBN: 0-07-008845-4] A practical guide to LaTeX for those who have never used TeX . Provides examples of what's needed

to do standard things without going into full details on how LaTeX works. Covers formatting environments, mathematical typesetting, tables, footnotes, bibliography, indices, and document styles. Concludes with a glossary and a display of various Computer Modern Fonts at different magnifications. LAS

Applications, P. *Transactions of the Seventh Army Conference on Applied Mathematics and Computing.* US Army Research Office (POB 12211, Research Triangle Park, NC 27709), 1990, xx + 704 pp, (P). Nearly 50 papers on a wide variety of topics in applied mathematics from a June 1989 conference at West Point. Image algorithms, inverse problems, game theory, wave propagation, tiling, splines, and much more. LAS

Applications, T, S, P, L*. *Mathematical Bioeconomics: The Optimal Management of Renewable Resources, Second Edition.* Colin W. Clark. Wiley, 1990, xiii + 386 pp, $49.95. [ISBN: 0-471-50883-7] A timely revision of the 1976 *First Edition* (TR, February 1977) emphasizing the "deliberate overexploitation" of biological resources possessing low rates of biological growth. "If forests, marine mammals, or grazing lands are incapable of replenishing themselves at sufficiently high rates, economically rational owners will tend to over-exploit these resources." Chapters feature economic models of resource management; optimal control; nonlinear, dynamical, stochastic, and discrete models; growth, aging, and multi-species models; and resource regulation. LAS

Applications (Biological Science), P, L. *Lecture Notes in Biomathematics-83: Mathematical and Statistical Approaches to AIDS Epidemiology.* Ed: C. Castillo-Chavez. Springer-Verlag, 1989, ix + 405 pp, $47.10 (P). [ISBN: 0-387-52174-7] Nineteen papers on modelling HIV infectivity, transmission dynamics, social dynamics of sexually transmitted diseases, and interactions with the immune system. BC

Applications (Biological Science), P. *Some Mathematical Questions in Biology: The Dynamics of Excitable Media.* Ed: Hans G. Othmer. Lect. on Math. in the Life Sci., V. 21. AMS, 1989, viii + 181 pp, $34 (P). [ISBN: 0-8218-1171-1] Six papers on excitable (as opposed to oscillatory) systems, including synchronization of cardiac pacemakers, collective behavior in the hippocampal, and wave propagation during aggregation in slime mold. BC

Applications (Communication Theory), P. *Signal Processing, Part I: Signal Processing Theory.* Eds: L. Auslander, T. Kailath, S. Mitter. IMA, V. 22. Springer-Verlag, 1990, xiii + 251 pp, $29. [ISBN: 0-387-97215-3] Seventeen papers from a 1988 workshop in signal processing. BC

Applications (Fluid Dynamics), P. *Computational Fluid Mechanics, Selected Papers.* Alexandre Joel Chorin. Academic Pr, 1989, xv + 223 pp, $35. [ISBN: 0-12-174070-6] Fourteen of the author's papers on the general themes of turbulence and the numerical solution of fluid mechanics equations with small viscosity. Originally published between 1967 and 1982, these papers are collected now because of the author's refreshing and reasonable belief that the ideas of the subject "are easier to learn if they are presented in the simpler garb that preceded the development of sophisticated implementations and of a general mathematical theory." JO

Applications (Fluid Dynamics), P. *Boundary Value Problems in Mechanics of Nonhomogeneous Fluids.* S.N. Antontsev, A.V. Kazhikhov, V.N. Monakhov. Stud. in Math. & Its Applic., V. 22. North-Holland (US Distr: Elsevier Science), 1990, xii + 309 pp, $85.25. [ISBN: 0-444-88382-7] An understandable but choppy English translation of the original 1983 Russian edition, this volume reports investigations into boundary value problems for systems of partial differential equations describing viscous gas flows and the filtration of multi-phase mixtures in porous media. Each chapter includes a section describing relevant open problems. JO

Applications (Physics), S(18), P. *Analysis and Continuum Mechanics.* Eds: S.S. Antman, *et al.* Springer-Verlag, 1989, x + 829 pp, $80.40 (P). [ISBN: 0-387-50917-8] A collection of papers dedicated to J. Serrin. MU

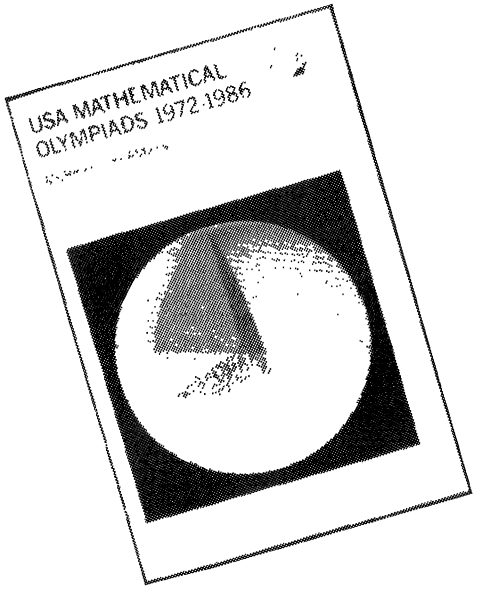
Applications (Physics), T(18), S, P. *Scattering Theory for Hyperbolic Operators.* Vesselin Petkov. Stud. in Math. & Its Applic., V. 21. Elsevier Science, 1989, xiv + 373 pp, $92.75. [ISBN: 0-444-88056-9] Applicable to spaces of either even or odd dimension, the approach used here involves the RAGE type theorem. Topics include contraction semigroups and power bounded operators, the Cauchy problem for the wave equations, scattering theory for symmetric systems with dissipative boundary conditions. MU

Applications (Simulation), P. *The Monte Carlo Method for Semiconductor Device Simulation.* C. Jacoboni, P. Lugli. Springer-Verlag, 1989, x + 356 pp, $98. [ISBN: 0-387-82110-4] It is getting to the (submicron) point where you cannot design a computer chip without first simulating the physics of charge transport. Monte Carlo (casino) methods are likely to help. BC

Reviewers

DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; SG: Steven Galovich, Carleton; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; RWJ: Roger W. Johnson, Carleton; JJ: Jason Jones, St. Olaf; LCL: Loren C. Larson, St. Olaf; JO: Jeff Ondich, St. Olaf; AO: Arnold Ostebee, St. Olaf; MLR: Margaret L. Reese, St. Olaf; KS: Karen Saxe, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; PZ: Paul Zorn, St. Olaf.

USA MATHEMATICAL OLYMPIADS



Every year 100 of the most mathematically talented high school students in the country compete in the USA Mathematical Olympiad (USAMO). The USAMO is the third stage of a three-tiered mathematical competition for high school students in the United States and Canada that begins with the AHSME taken by over 400,000 students, continues with the American Invitational Mathematics Exam involving 2,000 students, and culminates in the 100-contestant USAMO.

USA MATHEMATICAL OLYMPIADS 1972-1986, PROBLEMS AND SOLUTIONS

Compiled by Murray S. Klamkin

People delight in working on problems "because they are there, " for the sheer pleasure of meeting a challenge. This is a book full of such delights. In it, Murray S. Klamkin brings together 75 original USA Mathematical Olympiad (USAMO) problems for years 1972-1986, with many improvements, extensions, finger exercises, open problems, references and solutions, often showing alternative approaches. The problems are coded by subject and solutions are arranged by subject as an aid to those interested in a particular field. Contains a glossary of frequently used terms and theorems, and a comprehensive bibliography with items numbered and referred to in brackets in the text. The problems are intriguing and the solutions elegant and informative. Students and teachers will enjoy working these challenging problems. Indeed all those those who are mathematically inclined will find many delights and pleasant challenges in this book.

180 pp., 1988, ISBN-0-88385-634-4

List: $13.50 MAA Member: $12.50

Catalog Number NML-33

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

Surfaces. Vector Fields. Differential Operators. Integral Flows. Time Animation. On your PC or Macintosh.

Fields&Operators

Introductory price $59.95

From the creators of  **the Complex Variables Program.**

Lascaux Graphics 3220 Steuben Ave., Bronx, NY 10467 (212) 654-7429

New. . .

Writing Mathematics Well, by Leonard Gillman
64 pp., 1987, ISBN-0-88385-443-0
Catalog Number - WMW List: $6.00 MAA Member: $4.50

Good writing conveys more than the author originally had in mind, while poor writing conveys less. Well-written papers are more quickly accepted and put into print and more widely read and appreciated than poorly written ones—and for notes, monographs, and books the quality of writing is more important than it is for papers.

In **Writing Mathematics Well,** Leonard Gillman tells his readers how to develop a clear and effective style. All aspects of mathematical writing are covered, from general organization and choice of title, to the presentation of results, to fine points on using words and symbols, to revision, and finally, to the mechanics of putting your manuscript into print. No book can by itself make you a better writer, but this one will alert you to the opportunities for better and more forceful writing. It does this both by precept and by example.

A book to be read for its sharpness and wit as well as for enlightenment, **Writing Mathematics Well** should be on the shelf of anyone who writes or intends to write mathematics. It will amuse and delight the already careful writer and it will help reform and refine the sensibilities of those who may be somewhat careless about their writing.



Order from
The Mathematical Association of America
1529 Eighteenth St. NW
Washington, D.C. 20036

TWO BOOKS ON MATHEMATICS EDUCATION

TOWARD A LEAN AND LIVELY CALCULUS
Report of the Conference/Workshop to Develop Curriculum and Teaching Methods for Calculus at the College Level

Ronald G. Douglas, Editor
MAA Notes #6
249 pp., Paperbound, 1987,
ISBN-0-88385-056-7
List: $12.50
Catalog Number NTE-06

Should calculus be taught differently? Can it? Common wisdom says "no"—which topics are taught, and when, are dictated by the logic of the subject and by client departments. The surprising answer from a four-day Sloan Foundation-sponsored conference on calculus instruction, chaired by Ronald Douglas, is that significant change is possible, desirable and necessary. Meeting at Tulane University in New Orleans in January, 1986, a diverse and sometimes contentious group of twenty-five faculty, university, and foundation administrators, and scientists from client departments put aside their differences to call for a leaner, livelier, more contemporary course, more sharply focused on calculus's central ideas and on its role as the language of science.

This volume contains the results of that conference and the papers presented to the conferees. These are certain to be the point of departure and basis for efforts to strengthen and reshape calculus in the next decade.

CALCULUS FOR A NEW CENTURY: A PUMP NOT A FILTER

L.A. Steen, Editor
MAA Notes #8
272 pp., Paperbound, 1987,
ISBN-0-88385-058-3
List: $12.50
Catalog Number NTE-08

Proceedings of the Colloquium held in October 1987 in Washington, D.C. to discuss calculus reform. The mathematical community is challenged to make the introductory calculus course into a pump that feeds more students into science and engineering, not a filter that cuts down the flow. These proceedings, with contributions from over eighty authors, show the full sweep of concerns and approaches of all the groups involved in calculus reform, including those currently teaching traditional and innovative courses, those whose students or employees need to use calculus as a tool, and the department chairs, deans, and others who must mobilize the resources needed for this reform.



ORDER FORM

Please send the following books:

Quantity	Catalogue Number	Unit Price	Price
____	NTE-06	$12.50	_____
____	NTE-08	$12.50	_____

Total $ _____

Name: _____

Address: _____

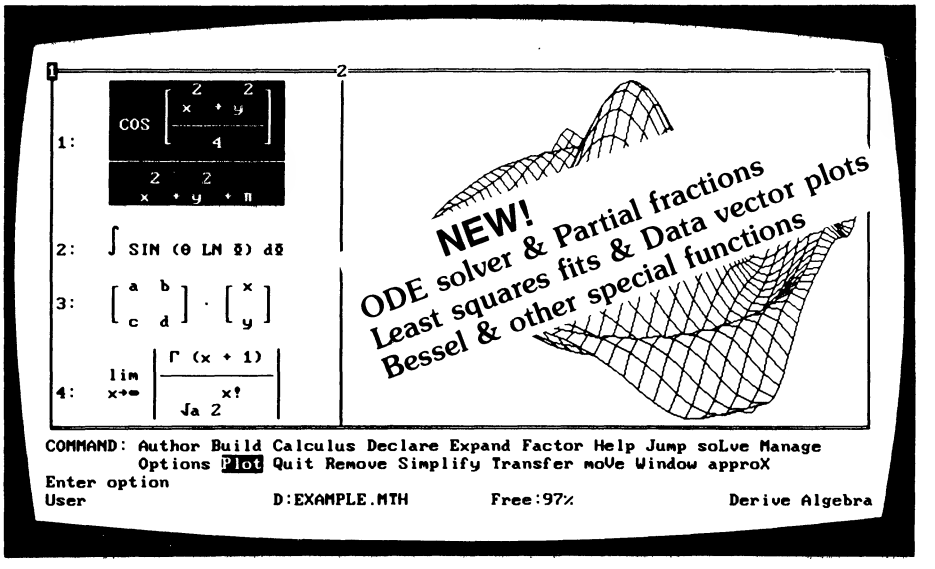
_____ ☐ Payment enclosed (sent postpaid)

_____ ☐ Please bill me (postage and handling extra)

Mail to: Mathematical Association of America, 1529 Eighteenth St., N.W., Washington, DC 20036

DERIVE

A Mathematical Assistant



2000 years of mathematical knowledge on a disk

DERIVE, the successor to **muMATH**, is a powerful computer algebra system for your PC compatible computer that provides the following capabilities:

- Exact and approximate arithmetic to thousands of digits
- Equations, complex numbers, trigonometry, calculus, vectors, and matrices
- 2D and 3D function plotting with zooming capability
- MDA, CGA, EGA, VGA, and Hercules graphics and text support
- Attractive 2D mathematical display of formulas
- Easy to use menu-driven interface with on-line help
- Ideal for engineers, scientists, students and teachers
- $200 plus shipping: **Call or write** for information.

System requirements: IBM PC or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (360K) or a 3½ inch (760K) diskette drive. **Or** NEC PC-9801 or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (640K) diskette drive.

DERIVE and muMATH are trademarks of Soft Warehouse, Inc. Hercules is a trademark of Hercules Computer Technology, Inc. IBM is a registered trademark of International Business Machines Corp. MS-DOS is a registered trademark of Microsoft Corp. NEC is a registered trademark of Nippon Electric Company.

Soft Warehouse INC
3615 Harding Avenue, Suite 505 • Honolulu, HI 96816
(808) 734-5801 after noon PST
Handcrafted software for the mind.

©1988 Soft Warehouse, Inc.

Winner of the First MAA Book Prize . . .



Carus Mathematical Monograph, #19

Field Theory and its Classical Problems,

by Charles R. Hadlock

323 + xvi pp. Hardbound

List: $28.00 MAA Member: $22.00

Field Theory and its Classical Problems is one of those rare monographs that will capture and hold a mathematically prepared reader (one reasonably familiar with basic calculus and linear algebra) until the last page. Mathematicians will be intrigued by the development which places Galois Theory in an historical perspective and allows it to unfold from the emergence of the classical construction problems to the discovery of their solutions. The book presents some of history's greatest and most elegant mathematics.

Read what some reviewers say about this monograph:

"Hadlock has produced a pedagogic masterpiece . . . His idea of beginning with the three Greek problems and letting them fire the reader's spark plugs is brilliant . . . (his) ability to inject vitality and enthusiasm into mathematical text is remarkable. (George Piranian, University of Michigan)

"In the preface, the author claims to have written the book for himself . . . Fortunately for us, he chose to share his work with the mathematical community. I suggested the book as collateral reading in a one-semester course in Galois Theory and the students found it very readable and most helpful in establishing a motivation and direction . . . This is a delightful book for both student and teacher." (John D. Leadley, ***Mathematical Reviews)***

Table of Contents

Chapter 1—The Three Greek Problems
Constructible Lengths; Doubling the Cube; Trisecting the Angle; Squaring the Circle; Polynomials and Their Roots; Symmetric Functions; the Transcendence of π.

Chapter 2—Field Extensions
Arithmetic of Polynomials; Simple, Multiple, and Finite Extensions; Geometric Constructions Revisited; Roots of Complex Numbers; Constructibility of Regular Polygons I; Congruences; Constructibility of Regular Polygons II.

Chapter 3—Solution by Radicals
Statement of the Problem; Automorphisms and Groups; The Group of an Extension; Two Fundamental Theorems, Galois' Theorem; Abel's Theorem; Some Solvable Equations.

Chapter 4—Polynomials With Symmetric Groups
Background Information; Hilbert's Irreducibility Theorem; Existence of Polynomials over Q with Group S_n.



Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Books that make mathematics fun for students to learn, and easy for you to teach . . .

Investigations in Algebra

Albert A. Cuoco

Investigations in Algebra departs from a preoccupation with calculus as the ultimate goal of, and the universal introduction to, advanced mathematics by using Logo to explore combinatorics, number-theory, the study of discrete functions, and other topics that are not on the traditional path to calculus. This approach encourages students to participate actively in exciting mathematics, developing their facility for abstraction as well as an appreciation for the power of mathematical methods.

"Back-to-basics enthusiasts would have us believe that experiential learning is incompatible with hard facts and measurable achievement. Al Cuoco, a gifted mathematician as well as a master teacher, gives his students a remarkable fund of solid, sophisticated mathematical knowledge, while respecting the exploratory process through which that knowledge was originally discovered. The book's use of computer programming as a medium for mathematical experiments helps make this work possible, without allowing the programming to displace the mathematics as the focus of attention." — Dr. Brian Harvey, Lecturer, Computer Science Division, University of California, Berkeley

350 pp. $29.95 paper, $45.00 cloth (August)

Approaching Precalculus Mathematics Discretely

Explorations in a Computer Environment

Philip G. Lewis

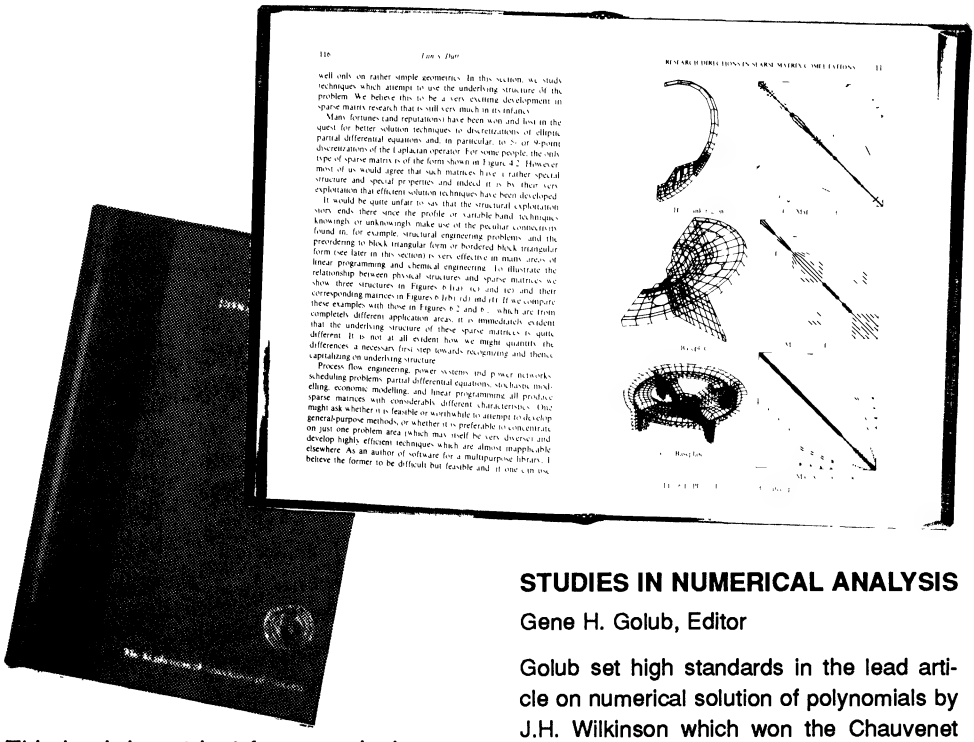
Lewis shows how concepts of discrete mathematics can be introduced through the computer using the Logo language, making them easier to teach and more fun to learn. He applies and explores much of the material in standard high school advanced algebra and precalculus programs. In conjunction with mathematical concepts, sophisticated programming techniques are developed, making this book a model for teachers looking for ways to integrate computers into the mathematics curriculum.

350 pp. $35.00 cloth, $19.95 paper (July)

To order IBM Logo™ call 1-800-IBM-2468
To order LCSI Logo II™ for Apple contact
Logo Computer Systems, Inc., 1-800-321-LOGO

The MIT Press
55 Hayward Street
Cambridge, MA 02142

STUDIES IN NUMERICAL ANALYSIS



This book is not just for numerical analysts, it is accessible to anyone with some mathematical foundation who wants to taste selections from a numerical analysis menu. The papers do a good job in motivating and introducing their subjects, and some provide a nice historical background. I recommend the articles in this book for anyone interested in learning more about the current research interests in numerical analysis. They may excite their readers to investigate further this relevant field of mathematics.

David C. Arney in *Mathematics and Computer Education*

STUDIES IN NUMERICAL ANALYSIS

Gene H. Golub, Editor

Golub set high standards in the lead article on numerical solution of polynomials by J.H. Wilkinson which won the Chauvenet Prize as the best piece of mathematical exposition to appear in 1985. These papers describe the wide range of research in numerical analysis. A variety of problems and of computational tools is treated. Some of the computations require only a hand-held calculator, while others require a powerful computer.

422 pp., 1985, ISBN 0-88385-126-1

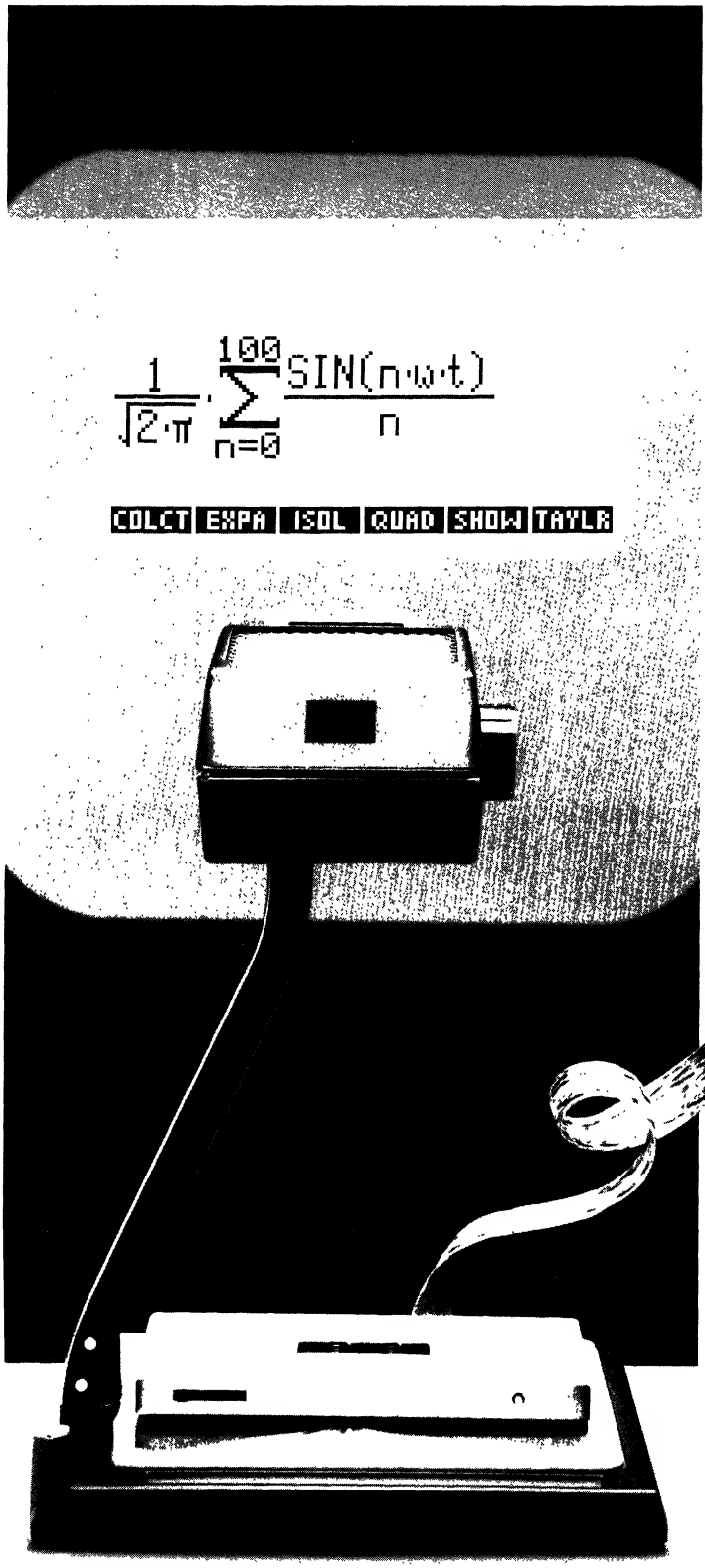
List: $46.50 MAA Member: $34.50

Catalog Number MAS-24

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

Help your students discover more meaningful relationships.



Free classroom display device with purchase of 30 calculators.

Showing is much more powerful than telling. So we've developed special classroom displays for our two most advanced calculators.

The new scientific expandable HP 48SX calculator and the cost-effective HP 28S are designed to put your students on the cutting edge of calculus and engineering. With more built-in functions and graphics solutions than any other calculator.

If your department or students purchase 30 HP 48SX or HP 28S calculators (or a mix of both), we'll give you a free HP 48SX and plug-in classroom display (an $850 retail value). Or a classroom display version of the HP 28S (a $500 retail value). And free teaching materials in the bargain.

Call us at **(503) 757-2004** between 8am and 3pm PDT for more details. Or write: Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330. Offer ends October 31, 1990, and applies to instructors at college and high school levels only.

There is a better way.

HEWLETT PACKARD

FOR YOUR LIBRARY



TWO-YEAR COLLEGE MATHEMATICS READINGS

Edited by Warren Page

viii + 304 pages. Paperbound.

List: $22.50 MAA Member: $16.50

This work surveys the two-year college mathematics landscape. Stops are made at points of interest that cover areas from algebra to recreational mathematics.

There are many things of interest for everyone in this book. Some are expository articles that introduce the reader to new advances in mathematics. Other articles can be used to enrich course work and to stimulate both the student's and instructor's mathematical intellect.

This book will serve as an excellent source of enrichment for the talented student. Teachers of undergraduates will find it is an excellent reference source that they will want to consult when preparing lectures.

Topics covered are:

- Algebra
- Geometry
- Number Theory
- Calculus
- Probability and Statistics
- Calculators and Computers
- Mathematics Education
- Recreational Mathematics



Order From:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 7 August–September 1990

Contents

(ISSN 0002-9890)

ARTICLES

Has Progress in Mathematics Slowed Down? PAUL R. HALMOS 561

Two Theorems, Two Sculptures, Two Posters HELAMAN FERGUSON 589

PROBLEMS AND SOLUTIONS

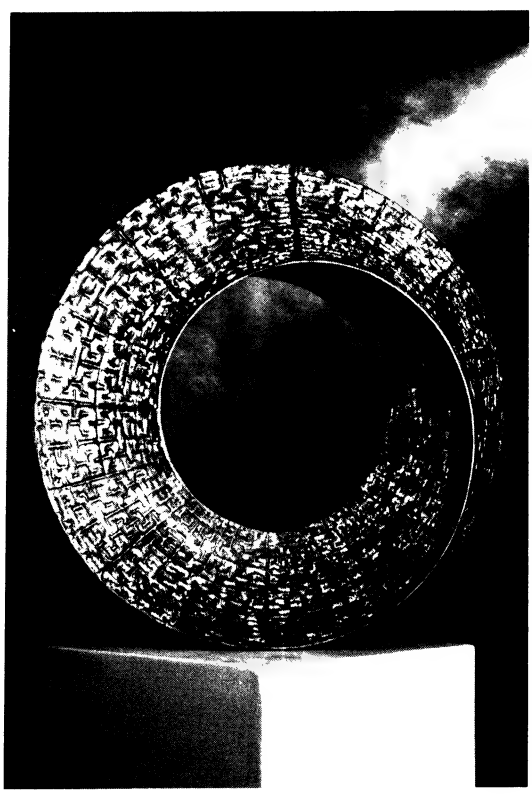
Elementary Problems and Solutions 611

Advanced Problems and Solutions 621

REVIEWS

Lattice Points, by Paul Erdös, Peter M. Gruber, and Joseph Hammer WILLIAM FIREY 633

TELEGRAPHIC REVIEWS 636



UMBILIC TORUS NC

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

Has Progress in Mathematics Slowed Down?*

PAUL R. HALMOS, *Santa Clara University*



PAUL HALMOS has three degrees from the University of Illinois; soon after getting the last one he became, for a couple of years, assistant to John von Neumann. Since then he has taught at many universities (including Chicago, Michigan, and Indiana) and has visited many others (including Miami, Montevideo, Hawaii, Edinburgh, and Western Australia); he has been on the faculty of Santa Clara University since 1985. His mathematical interests include ergodic theory, algebraic logic, and operators on Hilbert space.

Prologue. Do we know anything that Dedekind didn't know? We should. Dedekind died in February, 1916. Six weeks before that, late Friday afternoon on New Year's Eve 1915, the MAA was born in Columbus, Ohio. In connection with the celebration of its diamond anniversary, in August 1990, in Columbus, Ohio, it became my mission to report on whether and how mathematics has changed during the 75 years of the MAA's existence, and what follows is an attempt at such a report.

I am not trying to teach any mathematics in this report, nor even any history of mathematics—all I am trying to do is share an interesting look at the growth of mathematics in the last 75 years. Everybody could find out everything I found out by spending a few months looking at all extant volumes of *Mathematical Reviews* and a few dozen other journals, but everybody hasn't done it, and I have, and I am ready to tell what I have found.

The question that I set out to answer might be phrased this way: if you had a time machine to take you back to Dedekind, what could you teach him about progress in mathematics since his day? In an attempt to organize the possible answers, I propose to put them into three classes: concepts, explosions, and developments.

The first class consists of the new concepts that Dedekind could not have predicted or expected, the new words that all mathematicians alive today have heard over and over again and that many of us wish we knew more about. (Sample: catastrophe theory.) By an explosion I mean a piece of mathematical progress that is genuine mathematics, so recognized by the whole profession, but that is at the same time the answer to old problems of such great fame that it is hot news not only for the *Transactions*, but also for the *Times* for a day, for *Time* for a week, and for student mathematics clubs for many months. (Sample: the four-color theorem.) The third proposed class consists of the deep and in some cases even breathtaking developments (but not explosions) of the kind that may not make the *Times*, but could possibly get Fields medals for their discoverers. (Sample: the independence of the continuum hypothesis. Incidentally: not all the developers mentioned below got Fields medals, and not all Fields medalists of the period are mentioned below, but qualitatively the two classes are essentially the same.) Many (but not all) of the

*A greatly condensed version of this paper was presented in August 1990 at the Columbus meeting of the Mathematical Association of America.

subjects have been and some of the others no doubt soon will be treated in expository articles in the MONTHLY.

I am listing a total of 22 subjects: 9 concepts, 2 explosions, and 11 developments. If someone else had picked 22, the chances are that the overlap would have been considerable (10? 18?) and so would the difference. There is nothing wrong with that. Some of the subjects that could (should?) have been included but were not are: exotic spheres, the Hauptvermutung, NP completeness, pseudodifferential operators, and the simplex method. One thing is for sure: no matter what subjects were treated, not all would be of equal importance. Some of the topics that I discuss are yes-or-no theorems (such as the solution of Hilbert's fifth problem), some are tools (such as the fast Fourier transform), some are attitudes (such as nonstandard analysis), and some are gigantic theories (such as the Atiyah-Singer index equation). Each topic that is listed is here because either it is important or it is amusing or, at the very least, it has been well publicized, and, in any case, it deserves attention. All the topics do have at least one feature in common: none of them belongs to what is sometimes called applied mathematics and (in particular?) none of them belongs to computer science. I had two reasons for that decision: one is that I don't know the subject and the other is that it will surely be represented in other reports.

The articles vary in length from one paragraph to about a dozen. The length was determined, in each case, by how complicated the subject is, by what I was able to learn of it, and, of course, to some extent, by personal preference. It is unusual in research mathematics to find something that is both interesting and elementary; when I did find something like that, I took advantage of the find and digressed somewhat from the main point.

If one of the pieces of mathematics that I am reporting on is one that you know about, you won't learn anything from what I say, but if not, you are in danger of learning something. You will probably not learn the precise statement of a theorem (mathematics is, after all, not a collection of theorems but a collection of ideas), and you will certainly not learn a proof. The purpose of the articles is not so much to explain things as to put them into a context. That can mean many different things; one of them is that an article about a generalization of something should at least refer to the easiest nontrivial manifestation of that something. Another thing it means is that the articles are written in prose; they contain very few of the customary abbreviating symbols of mathematical exposition. In a recently much publicized phrase, the articles are intended to contribute to the mathematical version of "cultural literacy"; they don't say what a subject really is, but they say enough about it to make casually dropped tea-time references more comprehensible than they were before.

A bibliography for a report such as this would almost be a bibliography for all mathematics. As a compromise between that and nothing, I offer exactly one reference at the end of each section. The understanding of a reader who consults one of these references would, I hope, be increased by the explanation there. In addition, such a reader would, I know, find further references to the articles and books that treat the subject of interest.

The unique reference for each section is either to

The Encyclopedic Dictionary of Mathematics
(abbreviated EDM)

(prepared by the Mathematical Society of Japan, M.I.T. Press, 1980) or to the journal

The Mathematical Intelligencer

(abbreviated MI).

EDM references appear in the form

EDM 146/B

(referring to the article labelled 146/B), and MI references appear in the form

MI 8/1/40

(referring to volume 8, number 1, page 40).

That's it; now let's get down to work, and see what these concepts, explosions, and developments really are.

Concept 1: Moore-Smith limits. Modern general topology, and in particular its manifestations in the so-called weak topologies of certain function spaces, taught us that limits of sequences are no longer a strong enough tool for analysis. In classical analysis we learn that a set is closed if and only if it contains the limits of all convergent sequences in it, but there are many important and useful topological spaces for which that statement is not true. If, however, sequences are replaced by "generalized sequences" (the streamlined word is "nets"), and, correspondingly, ordinary limits of sequences are replaced by Moore-Smith limits of nets, the classical proofs work again, often with no changes except terminological ones, and they yield results just as useful as the classical ones. (Example: a set is closed if and only if it contains the Moore-Smith limits of all convergent nets in it.)

The Moore of the seminal 1922 Moore-Smith paper is the great E. H. Moore (one of the teachers of R. L. Moore) and the Smith is the otherwise largely forgotten H. L. Smith. The word "net" was first used, so far as I know, by J. L. Kelley in 1950.

Reference: EDM 89/H.

Concept 2: Distributions. The set $\mathbb{S}$ of infinitely smooth functions $\mathbb{R}^n$ with compact support is in an obvious way a vector space. Let us say that a sequence $\{\varphi_m\}$ of such functions converges to 0 if the supports of all the φ_m are covered by some fixed compact set, and not only does the sequence converge uniformly to 0, but, in fact, so do all its mixed partial derivative sequences.

If now f is a complex-valued integrable function on $\mathbb{R}^n$, then the equation

$$T_f(\varphi) = \int \varphi(x) f(x)\, dx$$

defines a continuous linear functional on $\mathbb{S}$, and the same is true if f is only locally integrable. If μ is a complex finite measure in $\mathbb{R}^n$, then the equation

$$T_\mu(\varphi) = \int \varphi(x)\, d\mu(x)$$

defines a continuous linear functional on $\mathbb{S}$, and the same is true if μ is required to be finite on compact sets only. If, for instance, $\mu(E) = 1$ when the origin belongs to E and $\mu(E) = 0$ otherwise (in other words, μ is a point mass at 0), then $T_\mu(\varphi) = \varphi(0)$ for every φ—which may remind you of the curious behavior of something that used to be called the Dirac delta function (but which, of course, never was a function).

All continuous linear functionals on $\mathcal{S}$ are called *distributions*; the inspiration of Fields medalist Laurent Schwartz (and the other giants on whose shoulders he stood) was to realize that the concept of distribution generalizes the concept of function and that in many classical situations, in which no function exists that satisfies certain prescribed conditions, a distribution satisfying them may exist and is for all practical purposes just as good. A typical kind of "prescribed condition" is a partial differential equation to be solved—a distribution solution often gives as much applicable information as a classical honest one.

Reference: EDM 130/B.

Concept 3: The Monte Carlo method. Mark a table with a system of parallel lines spaced, say, two inches apart from one another, and consider a needle (a mathematically idealized needle, in other words a line segment) of length one inch. Drop the needle on the table at random and ask: what is the probability that the needle will not land between a pair of the parallel lines but will cross one of them? This is the famous Buffon needle problem, and the method of working it out is not especially difficult; it is an exercise in many elementary probability texts. The answer turns out to be $1/\pi$.

Is that a bit of a surprise? How does π get into the answer to *this* question? That's not the main issue right now, but as long as it's been raised a comment on it might be in order. The answer to the probability question depends, of course, on how the needle gets on the table—or, to say the same thing more precisely, it depends on the probability distribution that is tacitly or otherwise assumed to be the one that the fall of the needle is subject to. Different possible distributions have been studied, and, to nobody's surprise, yielded different answers. One possibility is to let the probability distribution of the location of the center of the needle be uniform on a line segment perpendicular to the parallel lines and of length three inches, say, and, independently, to let the probability distribution of the angle at which the needle is inclined to that perpendicular segment be uniform between 0° and 180°. Once angles are explicitly mentioned that way, it should be no surprise that π enters the answer.

If the answer is accepted, a curious consequence follows. Perform the experiment, repeatedly, many times, and count the ratio of the number of successes to the number of trials—the law of large numbers tells us that that ratio will be very nearly equal to $1/\pi$. Conclusion: the value of π can be determined from a physical (probability) experiment, with no calculation (except the formation of a reciprocal at the end). This was known and appreciated a couple of hundred years ago—it was a forerunner of the modern technique known as the Monte Carlo method, introduced around 1945 by Ulam and von Neumann.

The Monte Carlo method has had frequent applications when the obvious calculations that a question demands seem formidable—replace the question by a probability question that has the same answer and find the answer to the probability question either by actual experimentation, or, more likely, by computer simulation of random experimentation. An easy example is the evaluation of a definite integral $\int_0^1 f(x)\,dx$ where f is a bounded function (for definiteness, $0 \leqq f(x) \leqq 1$ whenever $0 \leqq x \leqq 1$). Proceed as follows: choose (or let your computer choose for you) a large set of pairs (x, y) of random numbers, with $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$, and then evaluate the ratio of the number of them for which $y \leqq f(x)$ to the total

number. That ratio is approximately equal to the average $\int_0^1 f(x)\,dx$—and, lo and behold, the integral has been (approximately) evaluated.

This example is a little too naive for the theory, but the genuine real-life applications are in spirit pretty much the same. To calculate something too messy to calculate, whether or not the something is already a probability, replace the question by a probability question (or a different probability question) that is more easily susceptible of experimental (or, rather, computer imitated experimental) evaluation.

Why the name Monte Carlo? Presumably because that's the name associated with one of the most famous gambling places of the world. What the name suggests is that if you can't solve a problem, try gambling with it—that is, replace it by a gambling problem that has the same solution and that you have a better chance of being able to solve.

Reference: EDM 378/B.

Concept 4: Categories. If you understand vector spaces and linear transformations, you are 90% of the way toward understanding categories. Mathematicians have been viscerally aware for a long time that vector spaces and linear transformations behave "just like" groups and homomorphisms, and they in turn behave just like topological spaces and continuous mappings—the parallelisms are great and all pervasive. If you know what subgroups are, you know what subspaces are, and if you understand how to form quotient groups, you are in a good position to discover (to rediscover) how to define quotient spaces.

Category theory is a formalization of that visceral understanding; it was born (to Eilenberg and Mac Lane) in 1945. A category is a class of "things" of two kinds, objects and morphisms, satisfying three easy axioms. Brutally summarized, those axioms say that morphisms can be composed, provided only that their domains and ranges match properly, that for each object X there is an identity morphism from it to itself, and that if $\mathrm{Hom}(X, Y)$ denotes the set of all morphisms from X to Y, then the only way $\mathrm{Hom}(X, Y)$ and $\mathrm{Hom}(X', Y')$ can have any elements in common is to have $X = X'$ and $Y = Y'$.

Vector spaces, and groups, and topological spaces fit under this scheme, and so do abstract sets, and rings, and differentiable manifolds, and modules, etc., etc. More special concepts than just morphisms can be defined within each category (such as monomorphisms, subobjects, direct products, and duality); they are all pleasant and handy and none of them holds any surprises.

An important part of the theory of categories is the theory of functors. A functor is a way of associating with the objects and morphisms of one category the objects and morphisms of another. Typical example: each vector space has a dual, and each linear transformation between vector spaces has an adjoint. Trivial but helpful example (called the forgetful functor): each topological space determines a set and each continuous mapping determines a mapping—just forget the topology and the continuity, and thus pass from the category of topological spaces to the category of sets.

Category theory started as (and continues to be) a convenient language in which to describe many phenomena, and for many of us that's all it is. For the fascinated specialist it is a subject of research that can discuss quotient categories, adjoint

functors, categories of categories, and other such towers of elaborate complications of great fascination—long may it wave.

Reference: EDM 53/A.

Concept 5: *K*-theory. If $\mathbb{U}$ and $\mathbb{V}$ are finite-dimensional vector spaces with the field of real numbers acting as scalars, then so is the direct sum $\mathbb{U} \oplus \mathbb{V}$ of $\mathbb{U}$ and $\mathbb{V}$. At a casual glance the formation of direct sums looks like a perfectly respectable addition process, but a second glance shows that there is something a little wrong: the process isn't associative. The direct sum $(\mathbb{U} \oplus \mathbb{V}) \oplus \mathbb{W}$ resembles the direct sum $\mathbb{U} \oplus (\mathbb{V} \oplus \mathbb{W})$ in many ways, but strictly speaking it's not the same; the elements of the first are ordered pairs whose *first* coordinates are ordered pairs, whereas the elements of the second are ordered pairs whose *second* coordinates are ordered pairs. The temptation to consider the correspondence that assigns to every element $((u, v), w)$ in the first direct sum the element $(u, (v, w))$ in the second is well nigh irresistible. That correspondence is an isomorphism between the direct sum spaces involved, and the natural way out of the non-associative obstacle is to identify two vector spaces if they are isomorphic. If that is done, and that is the normal thing to do in this situation, another small obstacle arises, but it's easy to make it too go away: it has to be checked that the addition process can be unambiguously defined for isomorphism classes of vector spaces as well as for just plain vector spaces. That is, what has to be shown is that if $\mathbb{U}$ and $\mathbb{U}'$ are isomorphic vector spaces and if $\mathbb{V}$ and $\mathbb{V}'$ are isomorphic vector spaces, then the direct sum $\mathbb{U} \oplus \mathbb{V}$ is isomorphic to the direct sum $\mathbb{U}' \oplus \mathbb{V}'$. That is true and almost obvious, and once it is on record, then the direct sum of two isomorphism classes of vector spaces can be defined by choosing a representative element of each class, forming the direct sum of those representative elements, and then forming the isomorphism class of that direct sum.

Very well—a sort of addition can be defined for vector spaces (yes, yes—they should be called isomorphism classes of vector spaces, but everyday mathematical idiom continues to call them vector spaces anyway, and mentally remembers that equality no longer means honest equality but isomorphism)—what can be said about the class of all vector spaces under this addition? Natural question: do they form a group? The first part of the answer is easy—there is an identity element, that is, there is a neutral vector space with the property that adding it to any other produces no changes—namely, the unique 0-dimensional vector space. The next part of the answer is dishearteningly negative: no, vector spaces under direct sum do not form a group, because inverses, negatives, do not exist.

The setback is not serious—it is no more serious than that encountered by a child who has learned to add positive integers, has discovered 0, and is discouraged by his inability to go backward—subtraction doesn't always make sense. To subtract u, say, from $u + v$ is something that can be done, but subtraction of an arbitrary positive integer from another may or may not be doable. Similarly, someone might be willing to say that subtracting $\mathbb{U}$ from $\mathbb{U} \oplus \mathbb{V}$ is permissible, and yields the answer $\mathbb{V}$, but what about the general case, and, in particular, what about subtracting $\mathbb{U} \oplus \mathbb{V}$ from $\mathbb{U}$?

The way out, for vector spaces, is the same as the way out was for integers—the integers that do not exist are created by fiat, or by an explicit set-theoretic construction: negative numbers, and negative direct sums (!) are adjoined to the system under the study, and the result is a perfectly good group.

What group is it? The answer is easy to state and easy to understand—it is, except for notation, the same as the additive group of all integers. The point is that if isomorphic finite-dimensional vector spaces are identified, then there is nothing to distinguish between two vector spaces except their dimension—which is a nonnegative integer—and, consequently, the process of adjoining "negatives" to vector spaces (or, rather, to their isomorphism classes) is exactly the same as the process of adjoining negatives to the nonnegative integers.

The procedure indicated above can be carried out, and turns out to be profitable to carry out, even if the field of real numbers is replaced by an arbitrary ring. Rings occur in "the real world" more often than fields do, but they constitute a more difficult subject. Algebraic K-theory is a partial attempt to face the difficulties. What makes the use of fields easier is that a finite-dimensional vector space has a finite basis. No, that's not a redundancy. To see that it is not, rephrase it this way: a finitely generated module over a field has a finite basis. (Recall that the definition of a module is just like that of a vector space except that an arbitrary ring is allowed to play the role of the coefficient field.) The point is that the corresponding statement for modules over rings instead of fields is not always true. It *is* true for "good" modules, but a precise definition of "good" is a technicality that can safely be omitted from an overview such as this one. (According to a popular definition the good modules are the ones that are finitely generated and projective.)

All right then, if vector spaces are replaced by modules, the group theoretic construction goes pretty much the same way. Addition is defined for "good" modules by forming direct sums, and then the definition has to be modified, as before, so as to apply to equivalence classes of modules instead of modules; the result is a lovely associative addition, with a zero element, but no inverses. (Caution: the useful notion of equivalence here is a slightly weakened version of isomorphism, but the principal features of the theory remain the same anyway.) Let inverses, negatives, be adjoined, by fiat, or by the usual construction of forming ordered pairs ("differences") and equivalence classes thereof, and the result is a group. The equivalence relation that defines the group assigns to each module over a ring R an element of the group—that element plays the role of the generalized dimension of the module. The group thus associated with the prescribed ring R is denoted by $K_0(R)$ and is called the Grothendieck group of R—and that association is the first step of what is called K-theory.

K-theory does more than that first step: in general it associates an abelian group $K_n(R)$ with every ring R and every nonnegative integer n, in ways that become more and more mysterious as n increases. For $n = 0$, the result is a generalization of dimension; for $n = 1$, it is a generalization of determinant; after that—don't ask. But, in any event, it can't do any harm to take a quick look at $n = 1$.

The objects at the center of the stage this time are not modules, but automorphisms of modules. If the coefficient ring is a field, a good assumption to begin with, then automorphisms can be viewed as nonsingular matrices. When should two such objects be regarded as equivalent? One classically inspired possibility is to call two matrices equivalent just when they can be obtained from one another by a finite sequence of elementary transformations. Elementary transformation? In this context that means: add an arbitrary scalar multiple of any row (or column) to any other row (or column). Alternative definition: call a matrix elementary if it is either equal to the identity matrix or differs from it at exactly one non-diagonal

entry, and then define an elementary transformation as multiplication, fore or aft, by an elementary matrix. Since an elementary matrix obviously has determinant 1, it follows that every matrix in the group generated by the set of all elementary matrices has determinant 1, and that two equivalent matrices must have the same determinant. The two theorems in the preceding sentence are the weaker halves of two better theorems, which say that a matrix has determinant 1 *if and only if* it is a product of elementary matrices, and that two nonsingular matrices (over a field) are equivalent *if and only if* they have the same determinant. Summary: the quotient group of the group of all invertible matrices by the (normal) subgroup generated by the elementary matrices is equal (well, isomorphic) to the multiplicative group of nonzero elements of the coefficient field.

The point of this summary is that it is a nontrivial statement about determinants of matrices over fields in which the word "determinant" does not occur. Determinants of matrices over rings are hard to come by, but elementary transformations and elementary matrices make perfectly clear sense. The concept they lead to must cope with two kinds of matrix operations—the algebraic ones of sum and product and the geometric one of direct sum. The best way to do that is to apply similar considerations to infinite matrices each of which is in fact a direct sum of an infinite identity matrix and a finite invertible matrix (with entries in a prescribed ring R). The resulting quotient group (modulo the subgroup generated by the set of all elementary matrices) is denoted by $K_1(R)$ and is called the Whitehead group of R. It turns out to be abelian (a nontrivial theorem), and its elements are the generalizations of determinants (of matrices over fields) in the same sense in which the elements of $K_0(R)$ are the generalizations of dimensions (of vector spaces over fields).

The uses of K-theory in algebra and topology are many, and, in particular, the topological cousin of the algebraic theory discussed above is essentially involved in the proof of the Atiyah-Singer generalization of the Riemann-Roch theorem.

Reference: EDM 236/I.

Concept 6: The fast Fourier transform. The Fourier transform of a function f on the real line is usually defined to be the function g given by

$$g(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(y)\, e^{-ixy}\, dy,$$

assuming that the integral exists in some decent sense. The Fourier transform is a powerful tool in modern pure analysis and it is also a powerful tool in classical hard analysis and many of its concrete applications. In the applications it is, of course, desirable to minimize the computations that enter, and it is not surprising that various approximations are frequently introduced, such as replacing the infinite interval by a finite one and replacing the integral by a sum.

The discrete Fourier transform comes therefore to play a role—it is the transformation that assigns to an n-tuple $(x_0, \ldots, x_{n-1})$ of complex numbers the n-tuple $(y_0, \ldots, y_{n-1})$ given by

$$y_q = \frac{1}{n} \sum_{p=0}^{n-1} x_p\, e^{-2\pi ipq/n}, \qquad q = 0, 1, \ldots, n-1.$$

The so-called fast Fourier transform is a minor observation—an inspired minor

observation—whose discovery is variously attributed to J. W. Cooley and J W. Tukey (1965) and to C. F. Gauss (1805). The observation is that the sum, the discrete Fourier transform, can be evaluated by expressing it as a sum of clever subsums and making clever use of the algebraic properties of the roots of unity (the numbers $e^{-2\pi ipq/n}$) that enter. The popular name of those algebraic properties is "trigonometric identities". The idea that makes cleverness possible is that the Vandermonde matrix formed with the powers of a primitive nth root of unity can, for some n's, be factored in such a way that the factor matrices have many entries equal to 0, and that, as a result, the subsum evaluations need less arithmetic.

The evaluation of each y_q by the definition of the discrete Fourier transform involves the evaluation of $n - 1$ products (x_p times $e^{-2\pi ipq/n}$ for $p = 1, \ldots, n - 1$) followed by $n - 1$ additions (the partial sums). Since there are n values of y_q, the total work involves $n(n - 1)$ additions and the same number of multiplications.

The simplest version of the fast Fourier transform applies to the case in which n is represented as a product of two factors, $n = hk$. In that case the fast Fourier transform can involve as few as $n(h + k - 1)$ additions and $n(h + k)$ multiplications. If, for instance, $n = 100 = 10 \times 10$, then the direct method takes 9900 additions and 9900 multiplications, whereas the fast method takes 1900 additions and 2000 multiplications—a huge difference, which gets huger as n grows and the process is iterated.

It's a beautiful and useful part of what used to be called numerical analysis (before it came to be called computer science). Is it more than a trick? Does it teach us anything about pure analysis, about the group of roots of unity—does it have valuable analogues in other groups?

Reference: MI 7/3/49.

Concept 7: Nonstandard analysis. Leibniz used infinities and infinitesimals, but he admitted feeling slightly queasy about them. His successors banished such things from the mathematical heaven and worked with ε's and δ's instead. The modern theory of nonstandard analysis dredged the forbidden concepts up from the underworld and is trying to reinstate them at the right side of Cauchy's throne.

A possible attitude toward infinitesimals is to regard them as "ideal" elements, similar in that respect to the points at infinity in the projective plane. Once assumed, or defined, or constructed, the main tool in working with them is the so-called transfer principle, which says (roughly) that anything sayable in the appropriate formal language and true about the nonstandard universe remains true about the standard universe.

The basic standard universe consists (roughly) of "individuals" (Urelemente), and sets of them, and sets of sets, and sets of sets of sets, and so on ad infinitum. The nonstandard universe has many extra elements; they are (roughly) functions with values in the standard universe, or, slightly more precisely, classes of functions that have been identified according to a suitable equivalence relation. There is a sense in which the new elements are reminiscent of sequences. A constant sequence, whose constant value belongs to the standard universe, plays the role of that value, but there are sequences that are infinitely small (converge to 0) and others that are infinitely large (diverge to ∞), and in the democracy of the nonstandard universe they are all treated equally.

Consideration of such universes leads to ordered fields that are like the field of real numbers, and play the role of that field in the theory, but that are non-Archimedean—that is, they do not share with the real numbers the property that a small thing added to itself often enough becomes a large thing. For those who are comfortable with the use of formal languages, and believe that infinitesimals and infinities are intuitively clear and pleasant concepts, nonstandard analysis seems to be an efficient tool—such people can use that language to think in and, they say, they are led thereby to make otherwise elusive discoveries. Their greatest single victory so far (the only one that every proselyting article quotes) is the first proof (by Allen Bernstein and Abraham Robinson) of an invariant subspace theorem about certain special operators in Hilbert space—and, with considerable justice, they refuse to be fazed by the fact that soon after the first proof (nonstandard) other (completely standard) proofs came along.

The future of the subject is not yet clear to most non-believers—will it or won't it become an established part of mathematics—will it or won't it displace ε's and δ's?

Reference: EDM 274/E.

Concept 8: Catastrophes. Consider the graph of the parabola given by the equation $y^2 = x$ and consider the mapping $(x, y) \mapsto x$ (projection) from that graph to (the positive part of) the x-axis. With exactly one exception every point of the graph has a neighborhood that is homeomorphic to its image, which is an open interval on the line. The exception is, of course, the origin; in no neighborhood of the origin is the projection mapping one-to-one. This phenomenon is described by saying that the origin is a singularity of the (smooth) mapping from a 1-dimensional manifold (the graph) to another (the axis).

Consider the surface of a sphere, project it perpendicularly to a plane, and note that the local behavior of the projection varies from point to point. Each point of the northern hemisphere has a neighborhood in which the projection acts as a homeomorphism, and the same is true of the southern hemisphere, but at points of the equator the projection has a kind of singularity called a "fold." It looks like a fold, locally, doesn't it?: a small neighborhood of a point of the equator is, in effect, folded over on itself by the projection.

Consider next a suitable cubic surface (visualize it as a plane folded over a part of itself and then folded back again at a slightly different angle), project it perpendicularly to a plane, and note that this time the projection can misbehave in different ways. In the case of the sphere, any point of the plane that had an inverse image at all had either one or two inverse images; for the cubic surface that number can be 1, 2, or 3. The points for which it is 3 correspond to the regular points of the projection mapping, the points for which it is 2 correspond to a fold singularity, and the unique point for which it is 1 comes from a singularity called a "cusp". The reason for the terminology this time is that the "fold" points on the surface project onto a semicubical parabola in the plane whose cusp, in the usual sense of the word, comes from the cusp singularity.

The mathematics of catastrophe theory studies the problem of classifying the singularities of smooth mappings (such as the projections above) between smooth manifolds (such as the parabola, the sphere, and the cubic surface above). A typical result (for the 2-dimensional case) is that every smooth mapping of a surface to a plane has, after possibly an appropriate small perturbation, nothing

but cusps and folds for its singularities. In higher dimensions there are similar (and surprisingly short) complete lists of "elementary catastrophes". The theory goes back to the work of Whitney in 1955; since then it has been extensively cultivated and applied by several mathematicians. The list of names includes Thom in France and Arnold in Russia; both of them have made important contributions to both the theory and its applications. Legitimate applications exist, for instance, to wave propagation problems, elastic stability, and geometric optics.

Where does the term "catastrophe" come from? The answer seems to be that it was suggested (by Thom?) as an appropriate way to describe startlingly discontinuous changes in the world around us. Here is a simple example: suspend a weight from a horizontal bar, and ask for the maximum supportable load. Intuition and experience suggest that as the load is increased nothing much happens, except possibly a gentle bend, till, suddenly, the bar breaks and the weight falls—a catastrophe. A more complicated example that is mentioned by almost everybody who writes on the subject is the degree of aggressiveness of a dog that is approached by a possible enemy. The dog feels both fear and rage—which is greater? What will the dog's response be: flight or fight? Once again intuition and experience suggest that the dog will cower at first and back away, till suddenly, because the distance to the enemy has decreased too much and the threat has increased beyond the supportable point, the dog will snarl and attack—a catastrophe.

How does it happen that a deep mathematical theory, with genuine and valuable applications, becomes as controversial as catastrophe theory has become? There is nothing wrong with the theorems that Whitney, Thom, Arnold, and others have proved; what then can a controversy be about? The answer is that some of the proponents of the theory (notably Thom himself, and Zeeman) are describing and claiming applications that seem to others (notably Arnold) far-fetched and unsound. The applications claimed are to economics, embryology, linguistics, psychology, and other subjects, and include political elections, mental disorders, prison riots, heart beats, stock market crashes, and the outbreak of war.

Here are four of the many comments that Arnold makes.

"The mathematical articles of the founder of catastrophe theory, René Thom, were reprinted as a pocket book—something that had not happened in mathematics since the introduction of cybernetics from which catastrophe theory derived many of its advertising techniques."

"A particular aspect of Thom's work on catastrophe theory is his original style: he established a fashion in not giving even sketchy formulations of results, let alone proofs."

"When the mapping we are concerned with is known in detail we have a more or less direct application of the mathematical singularity theory to various natural phenomena... In the majority of works on catastrophe theory, however, more controversial situations are considered where not only are the details of the mapping not known, but its very existence is problematical."

"The deficiencies of... catastrophe theory are too obvious to discuss in detail. We remark only that articles on catastrophe theory are sharply distinguished by the catastrophic lowering of the level of the requirement of rigour and also of the requirement of the novelty of the published results. Although one can understand the catastrophe theorists' reaction against the traditional flow in mathematics of rigorous but dull epigonic works, nevertheless their lack of respect to their

predecessors (to whom belong the majority of concrete results) can hardly be justified."

Reference: EDM 410/K.

Concept 9: Chaos. To a mathematician the word "point" is virtually synonymous with "element of a set". Let me give a moderately complicated set whose "points" are quite a bit more complicated than the points that most students usually meet. Begin with the closed unit interval $I = [0, 1]$, and let X be the set of all finite disjoint unions of nondegenerate closed subintervals of I. ("Nondegenerate" means that single points are not to be considered as intervals.) A "point" in the following paragraph is to be an arbitrary element of X.

Let T_C be the mapping of the set X into itself that acts on each point P by removing the open middle third of each of the closed intervals whose union P is. Thus, for instance, since I itself is a point of X, the image $T_C(I)$ makes sense and is equal to the set (point) $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. What happens when the mapping T_C of X into itself is iterated? Start, for instance, with I, and form successively $T_C(I), T_C^2(I), T_C^3(I), \ldots$. That decreasing sequence of finite unions of closed intervals is familiar to anyone who has ever taken a course in the set theory of real numbers: the intersection of that sequence of sets is the classical Cantor set. (The suffix C is intended to be a reminder of Cantor.)

The transformation so described is an artificial but interesting and suggestive example of a dynamical system. According to the most general definition, a dynamical system is a mapping of a set into itself. That definition, however, is too general to be of much use. The concept becomes useful when the underlying set and the transformation that acts on it have some structure of mathematical interest (algebraic, analytic, or geometric), and in all extant studies of dynamical systems such extra structures are indeed present.

The concept of "chaos" depends on that of dynamical system; roughly speaking, the theory of chaos is the study of the behavior (misbehavior?) of dynamical systems at infinity. Take any transformation T of any set X into itself, form the successive iterates $T, T^2, T^3, \ldots$, and then ask an intelligent question that involves them. Example: let X be the real line, define Tx to be $\cos x$ for every x in X, and ask: what happens to the sequence $x, Tx, T^2x, T^3x, \ldots$ for various values of x? The answer is easy to derive, but if you have never done it, I would suggest that first you take a hand-held calculator that has a "cos" button, start with any original input x that you like, and keep pushing "cos" over and over again. (The result looks a little more interesting if you use radians instead of degrees.)

Another example is the dynamical system T_C above. For still another example, much deeper, let X be the closed unit interval with its end points identified (so that this time a "point" is a real number between 0 and 1 inclusive, except that 0 and 1 count as the same point), and let T_2 be the act of doubling modulo 1. Thus, for example,

$$T_2\left(\frac{1}{3}\right) = \frac{2}{3},$$

$$T_2\left(\frac{3}{4}\right) = \frac{3}{2} \bmod 1 = \frac{3}{2} - 1 = \frac{1}{2},$$

$$T_2\left(\frac{1}{2}\right) = 0,$$

and

$$T_2(\sqrt{2} - 1) = T_2(1.4142\ldots - 1) = 2\sqrt{2} - 2 \,(\mathrm{mod}\, 1)$$
$$= 2.8284\ldots - 2 = 0.8284\ldots .$$

What happens to the sequence $x, T_2x, T_2^2x, T_2^3x, \ldots$ for various values of x? Can it converge to some limit? Can its closure contain an open interval? This example, relatively innocent as it may look, is in fact quite deep; it is a close relative of one that is topologically important (the so-called Smale horseshoe) and one that is analytically important (the so-called unilateral shift).

An example of a dynamical system that is more reminiscent of classical analysis than the ones above is obtained as follows: choose a couple of constants a and b, and define $T = T_{ab}$ in $\mathbb{R}^2$ by

$$T_{ab}(x, y) = (y + 1 - ax^2, bx).$$

Transformations of this kind were introduced and studied by Hénon. The formula defining them is not enough to scare even a student of elementary analytic geometry, but the study of their asymptotic properties is delicate and interesting. Those properties depend, of course, on the parameters a and b. If, for instance, $a = 1.3$ and $b = 0.3$, then there exists a set of seven points in the plane that constitute a "periodic attractor". That means that the iterated images of a point in the plane will either tend to infinity or else tend to get closer and closer to one of the seven points, and then to a next one, and then a next, and then, at the eighth step closer still to the first, and then to the second, etc. If, on the other hand, $a = 1.4$ and $b = 0.3$, then the iterated images of some points tend to infinity, whereas for other points those images cluster at a complicated set of curves that constitute what is called a "strange attractor" (the "Hénon attractor"). A good definition of that concept is yet to be offered. The main property of a strange attractor seems to be that it is an infinite set with "sensitive dependence" on the initial conditions (whatever that means). Strange attractors occur in, for instance, the study of turbulence.

Why is the word "chaos" used? The reason seems to be similar to the reason that caused catastrophes to be called catastrophes—it seems to be a subjective (not really a mathematical) reaction to an unexpected appearance of discontinuity. A possible source of confusion is that the startling discontinuity can occur at two different parts of the theory. Frequently a dynamical system depends on some parameters (the way the Hénon T_{ab} depends on a and b), and, of course, the sequence of iterates $x, Tx, T^2x, T^3x, \ldots$ of a dynamical system T applied to any particular x always depends on the initial point x. The startling change of the Hénon family (from periodic attractor to strange attractor) is regarded as chaos—unpredictability—and the very existence of the Hénon strange attractor, not obviously visible in the definition of the dynamical system, is regarded as chaos —unpredictability. I would like to register a protest vote against the attitude that the terminology implies. The results of nontrivial mathematics are often startling, and when infinity is involved they are even more likely to be so. It's not easy to tell by looking at a transformation what its infinite iterates will do—but just because different inputs sometimes produce discontinuously different outputs doesn't justify describing them as chaotic.

A part of the theory of chaos has acquired a kind of popularity (notoriety?), and no discussion of the subject can omit at least mentioning that part. I refer to the things called "fractals".

The artificial dynamical system T_C with which this exposition began led to the Cantor set, which occurs in chaos theory frequently and non-artificially. A curious and important property of the Cantor set is that its part in the interval $[0, \frac{1}{3}]$ is similar to the entire set—just blow it up by the dilatation $x \mapsto 3x$ and the small set becomes identical with the large one. Similarly, the part of the Cantor set in the interval $[\frac{2}{9}, \frac{7}{27}]$ is similar to the entire set; in fact, every neighborhood of every point of the Cantor set includes a subset that is similar to the entire Cantor set. The iteration of transformations tends to produce sets that exhibit such "self similarity", that is, sets that are invariant under change of scale; the Hénon attractor itself is another example.

The phenomenon is not new to the mathematical world. It has long been known that the Cantor set is the prototype of a large family of sets obtained by various architectural changes. Example: omit middle fifths, say, instead of middle thirds. Another example: replace the discarded open middles by, say, equilateral triangles. Still another: perform middle omissions not on an interval but on rectangles or disks or other interesting sets in the plane. All such sets exhibit local properties that are pleasing and surprising when first encountered but that become, upon acquaintance, so nonsurprising as to be almost trite—they are like friends whose company is as desirable as ever but whose shape and whose gestures remain completely familiar even when they grow old or wear disguises.

The Cantor set is a fractal, and so are various variations on that theme. (No definition of the general concept is needed here, but I mention in passing that a standard one is this: a fractal is a set whose dimension in the usual sense of topology, which is always less than or equal to its Hausdorff-Besicovitch dimension, is in fact strictly less. The topological dimension of the Cantor set is 0, and its Hausdorff-Besicovitch dimension is $\frac{\log 2}{\log 3}$.) The reason for the popularity of fractals is that they can be pretty. Certain planar versions of the Cantor set, with or without "colorization", look good on your living room wall, and disguising them by changing their sizes and shapes can produce enough pretty pictures to fill the whole wall or even an entire room in a museum of modern art.

"Art" is perhaps the operative word. A great mathematician conceives new theorems, discovers their beautiful proofs, and shares the results with the rest of us by publishing them. A great cook conceives new dishes, discovers their ingenious recipes, and shares the results with the rest of us by preparing them. An infinite Andy Warhol iteration of the label on a can of soup may be popular and it may be art, but it has very little to do with the genius of a cook. The infinite iteration of Cantor-like omissions may be a popular subject of conversation and it may be art, but it has very little to do with the insight of a mathematician. Paintings of cans of soup are not cooking, and paintings of Cantor sets are not mathematics.

Reference: MI 2/3/126.

Explosion 1: The four-color theorem. Everybody knows what the four-color problem is, but I'd like to begin by asking some elementary questions about it anyway, some questions that you may not have bothered to ask yourself before. The point is that the problem concerns the Euclidean plane, and the topology of the plane can exhibit some unexpected and weird phenomena.

Did you know, for instance, that there exist five regions in the plane ("the lakes of Wada") all of which have the same frontier? With "two" instead of "five"

everybody does know it: just remember the right and left half planes, that is, in terms of Cartesian coordinates, the set of points with positive x coordinates and the set of points with negative x coordinates. The common frontier is the y axis, and there is nothing surprising about that. (By the way, "region" in this context has its meaning familiar from a first course in complex variable theory: an open connected set.) What about "three" regions with a common frontier—is there a simple example of that? No, there is not; anybody who can construct an example with three can construct one with five just as easily, or, for that matter, with any other finite number.

Doesn't that sound like a negative solution of the four-color problem? The five regions represent five countries on a planar map, and since each two of them touch (at their common frontier), the only way to color the map correctly is to use five different colors.

Yes, that is a negative solution of the four-color problem, but only if the problem is carelessly formulated. If five regions have a common frontier, then they and their frontier must twist and interlock in an extraordinarily nasty manner. The classical four-color problem assumes (usually tacitly) that the countries represented on the map are healthy, not pathological. A simple way to rule out pathology is to demand that the frontiers of the countries be polygons (a finite number of line segments); the full depth and difficulty of the problem is present even in that strongly restricted case. If that seems too severe, it is all right to allow the frontiers to be pleasant continuous curves, but, in any event, as the five-region example shows, some geometric restriction is necessary.

A different kind of instructive example to look at is a map consisting of a circle in which a certain number of diameters have been drawn (say, for instance, three). Those diameters divide the circle into six countries each the shape of a slice of pie, and all six "touch" at the center. Does that mean that six colors are needed to color the map? No, it does not. That's a situation that the usual statement of the four-color problem does foresee and does rule out: two countries are said to "touch" only if they have a reasonably long piece of their frontiers in common, more than just some isolated corners.

Very well then: what is the smallest number of colors that is sufficient to color all decent maps? Could it possibly be two? One way to answer that question is to ask whether it is possible to design a map that represents three countries each of which touches both the others. The answer to that is easy: just make a circular map and make the three countries like three slices of pie with each of them having an angle of, say, 120° at the center: in that case each of them is "between" the other two. Conclusion: two colors are not enough to color all maps.

All right: is there a map with four countries so that each of them touches all three of the others? Yes—take the pie map just constructed and punch a hole in it —replace the center by a substantial circular country, with, say, half the radius of the entire map. The outside countries still have the property that each of them is between the other two, and, at the same time, each of them touches the center country. Conclusion: three colors are not enough to color all maps.

Does the technique used twice by now work for the next stage: is there a map with five countries each of which touches all others? In an attempt to see the answer, change the four-country map just studied in two ways: put the circle that it consists of (call it circle C) inside a larger circle (call it D), and cut the small circle in its center (call it B) into two parts. The resulting map has six regions: one is the

ring shaped region between D and C, two others are the two halves of B, and the other three were present before the changes. Call the five regions inside C five countries. A glance at the picture shows that three of them touch each of the other four, but two of them fail to do so. Adjoin the outside ring-shaped region (as a colony?) to that half of B that fails to touch each of the four countries—and, lo!, the result is a map of five countries every one of which touches all four others. A proper coloring of this map needs five colors—four are not enough. The technique is not difficult to generalize; for each positive integer n there exists a map of n countries each of which touches all $n - 1$ others.

Is something wrong? A possible objection is that (in the case $n = 5$) one of the five countries in the picture consists of two pieces—the country it represents is not connected. Is that wrong? Does that happen in the real world? Sure it does; the state of Michigan, for one, consists of two pieces. Does that mean that four colors are not enough for all maps? Yes, it does—provided that maps are allowed to contain disconnected countries. Once again the usual statement of the four-color problem foresees this obstacle, and rules it out; the problem concerns connected countries only.

If that five-color example is not acceptable, what would have to be done to prove that four colors are not enough to color all maps? An obvious possibility that the preceding discussion suggests is to try again to design a map with five countries, but this time connected ones, so that each of them touches all four of the others. Augustus de Morgan, in 1852, used ingenious topological arguments to prove that there is no such map (at least they would be called topological in the twentieth century)—but he was not allowed to jump to the conclusion that he had proved the four-color theorem. There was, perhaps, some temptation to jump to that conclusion, but it wouldn't have been right.

The point is this: to say that n colors are not enough to color a map is not the same as saying that it contains a submap of $n + 1$ countries each of which touches all n of the others—not at all. To see that, even for the already known case of $n = 3$, begin by contemplating a circular map with five pie-slice countries (each making an angle of 72° at the center). That map can be colored with three colors: just go A, B, C, A, B around the circle. Now punch a hole in the middle, that is, replace the center by a substantial circular country. The result is a map in which no submap of four countries has the property that each of them touches all three of the others—but, nevertheless, the entire map cannot be colored with only three colors. This argument proves again that three colors are not always enough, but this time the proof does not use four mutually touching countries. That is: many touching countries imply the need for many colors, but not conversely—many colors might be necessary for more subtle combinatorial reasons, and that's one of the sources of difficulty with the four-color problem.

In 1976 Appel and Haken offered a proof of the four-color theorem. Their main achievement was to reduce the problem to a finite one: they showed that once the answer is known for a finite list of specific maps (a very large list, to be sure), then it is known for all. That reduction is the first step of their proof; the second step is a computation. The reduction is standard mathematical reasoning—in principle (though definitely not in detail) it was known long ago. That is: it was known what kind of reduction was necessary, and many of its steps had already been taken. The computation was then carried out electronically, and, when the systematic exhaus-

tion procedure (which needed over 1000 hours of computer time on the first run through) encountered no obstacles, the computer ran up a flag and blew a fanfare of trumpets. Victory: the answer is yes, four colors suffice.

After the excitement died down, several grumbles and more severe complaints began to be heard: the program was never checked, some people said that it was uncheckable, and there were even dark rumors of errors, plain old-fashioned mathematical errors, gaps in the reduction proof. Grumbles and rumors such as that may be interesting, but right now they are beside the point. Assuming that the reduction is correct, assuming that the program is correct, and assuming that the electronic components functioned completely correctly (and they always do, don't they?), I want to ask: what did we learn from the proof? What do we know now that we didn't know before?

I do not find that an easy question to answer. To be sure: I am not going to spend my time looking for a counterexample to the four-color assertion. The printout had at least that practical effect: it discouraged attempts to prove it wrong. Except for that, however, I feel that we, humanity, learned mighty little from the proof; I am almost tempted to say that as mathematicians we learned nothing at all. Oracles are not helpful mathematical tools.

The most celebrated and difficult outstanding problem of mathematics is the Riemann hypothesis. There are many important theorems, with correct proofs on record, that are of the form "if the Riemann hypothesis is true, then" It would be good to know a proof (or disproof?) of the Riemann hypothesis, but till one is discovered the theorems in which it appears as a hypothesis can be illuminating and useful. That value judgment, however, does not change my opinion of oracles. If an oracle told me that the Riemann hypothesis is true, I don't think that my soul would be any richer for that single syllable.

The development of mathematics shortens proofs, gives insight, and deepens understanding by the discovery of ever new concepts and by the resulting subsumption of old ones under a suitably general theory that took years, decades, and sometimes centuries of labor to construct. Technical terms (Banach space, Artinian ring, fiber bundle) may frighten and alienate non-specialists, but they are just short phrases that abbreviate and clarify the effort, the work, and the thought of our great predecessors.

We may still be far from finding a "good" proof of the four-color theorem. We need a simple insight into a new and complicated kind of geometry or intricate algebra, and the distance from there to a purely conceptual, existential proof of the four-color theorem is probably just as great. I hold as an article of faith, however, that we have seen and travelled greater distances than that; the distance from the Peano axioms to the Atiyah-Singer index theorem could, for instance, be one of them. I believe that the computer (and, for another example, the 10,000 pages of published proof solving the simple groups problem) missed the right concept and the right approach. Their time will come. A hundred years from now both theorems (maps and groups) will be exercises in first-year graduate courses, provable in a couple of pages by means of the appropriate concepts, which will be completely familiar by then. Down with oracles, I say—they are of no use in mathematics.

I cannot, however, stop with that dictum—in all honesty, I should report that Appel and Haken do not completely share my religion. Speaking of their work,

they say: "We now know that a proof can be found. But we do not yet (and may never) know whether there is any proof that is elegant, concise, and completely verifiable by a (human) mathematical mind."

One final word. By an explosion I mean a loud noise, an unexpected and exciting announcement, but not necessarily a good thing. Some explosions open new territories and promise great future developments; others close a subject and seem to lead nowhere. The Mordell conjecture, the next explosion below, is of the first kind; the four-color theorem of the second.

Reference: EDM 165.

Explosion 2: Mordell's conjecture. Every grade school child knows that $9 + 16 = 25$, or, in other words, that $3^2 + 4^2 = 5^2$. A microsecond of thought reveals that the equation continues to hold when 3, 4, and 5 are replaced by $3k$, $4k$, and $5k$, no matter what k is. If the point of view is to look for integer solutions of the equation $x^2 + y^2 = z^2$, then there is really no virtue in distinguishing between $3, 4, 5$ and $6, 8, 10$. An efficient way of saying the same thing is to divide the equation through by z^2, so that it becomes $\left(\frac{x}{z}\right)^2 + \left(\frac{y}{z}\right)^2 = 1$, and then replacing $\frac{x}{z}$ and $\frac{y}{z}$ by x and y, so that the equation becomes $x^2 + y^2 = 1$, with the understanding that the solutions sought are rational numbers, not necessarily integers. From that point of view all the positive multiples of the $3, 4, 5$ solution become just one rational solution.

The equation $x^2 + y^2 = 1$ has many rational solutions, infinitely many—or, in geometric language, the curve that the equation defines has an infinite number of rational points on it. Those points are related to the so-called Pythagorean triples, and they are quite easy to find. The situation is different for the Fermat equations of higher degree; in fact, the celebrated infamous Fermat problem is to show that if $n > 2$, then the curve defined by the equation $x^n + y^n = 1$ has no rational points on it at all (except the trivial ones for which either x or y is 0).

The Fermat problem belongs to a broader context suggested by Louis Mordell in 1922. Mordell's genius in formulating the conjecture was to make the Fermat problem simultaneously harder and easier. Harder: enlarge the set of curves considered (so as to include all algebraic curves of genus greater than 1 over the field of rational numbers—a description that includes the Fermat curves of degree greater than 3). Easier: relax the conclusion to be drawn (so as to allow the existence of rational points, but, to be sure, only a finite number of them) Mordell's conjecture is a conjecture no longer—it was proved in 1983 by Gerd Faltings. Consequence: Fermat's problem may still turn out to have unexpected solutions, but, in any event, for each exponent it has only finitely many!

Reference: MI 6/2/41

Development 1: Ergodic theory. If you shoot a billiard ball perpendicularly into a side of the table, it will bounce straight back to the opposite side and then keep on bouncing back and forth between those two sides. If instead you shoot it straight from the middle of one side to the middle of an adjacent side, then it will bounce out at the same angle as it came in, hit the middle of the side opposite the one you started with, bounce onto the middle of the fourth side, and keep on going cyclically around like that. You can probably think of many other such

patterns—a carefully specialized original aim will produce a periodic pattern. What, however, happens if you shoot the ball at an angle that stands in an irrational relation to the sides of the table, or, more clearly put, if you shoot the ball at an angle that produces a nonperiodic orbit? Boltzmann, one of the founders of ergodic theory in the preceding century, thought about such things and formulated, among other things, what has come to be called the ergodic hypothesis. According to a primitive version of that hypothesis, the path of the (irrationally aimed) billiard ball will eventually pass through every point on the billiard table.

A little cogitation about the topology of a rectangle in the plane versus the topology of a curve in that rectangle might convince you that that version of the ergodic hypothesis is unlikely at the very least—the applicable laws of mechanics are not likely to produce plane filling curves. A modified version of the ergodic hypothesis came along: it was the conjecture that the path of the billiard ball, while it may not go through every point, will in any event get as near as you like to every point—that, in other words, it will be everywhere dense in the billiard table.

Statistical mechanics occupied itself with meditations such as these till G. D. Birkhoff came along with his seminal paper in 1931. What Birkhoff proved was that if you take any decent subset of the table, such as the right half, or the interior of a circle whose center is the center of the table—whatever—the average time that the billiard ball will spend in that subset, no matter where it started, is proportional to the area of the subset. The technically difficult part of his achievement is that it makes sense to speak of the average time—in other words that the limit involved in the meaning of that phrase actually exists—but the striking part of the conclusion is that the average time is a constant, independent of where the ball started. That conclusion implies, among other things, that if you look at the interior of any circle drawn on the table, no matter where it is drawn and no matter how large or how small it is, the billiard ball is certain to enter that interior over and over again—and hence, in particular, that the path of the ball is indeed everywhere dense.

Birkhoff's work was published well over 50 years ago, and since then ergodic theory has become a major part of mathematics that has rich connections with both classical and modern analysis and that is definitely one of the major steps forward that was taken in the last 75 years. (Family comment: G. D. Birkhoff's son is Garrett Birkhoff, both at Harvard in their days.)

Reference: EDM 146/B.

Development 2: Transcendental numbers. Is the number $2^{\sqrt{2}}$ rational? If not, is it at least algebraic (that is, a solution of a polynomial equation with integer coefficients)? The seventh of Hilbert's famous problems was about that question and the general context to which it belongs. A good way to formulate the general question is to ask when numbers of the form α^{β} are transcendental. If $\alpha = 0$, the question degenerates, and the same is true if $\alpha = 1$; similarly if β is either 0 or 1, the question is not one anybody wants to ask. If, more generally, β is rational, then the question transforms into easy and generally known subquestions concerning algebraic numbers. On the other hand, if either α or β is transcendental, then the answer sought for appears to be too close to the data. In view of these comments, the "right" question to ask is this: if α and β are algebraic numbers, with α different from both 0 and 1, and β irrational, does it follow that α^{β} is transcen-

dental? The answer turns out to be yes; it was obtained, more or less simultaneously and independently, by A. O. Gelfond ($\neq$ I. M. Gelfand) and Theodor Schneider in 1934. The question about $2^{\sqrt{2}}$ is covered: since $2 \neq 0$, $2 \neq 1$, and $\sqrt{2}$ is irrational, it follows that $2^{\sqrt{2}}$ is transcendental.

The theory didn't stop there. Here, for instance, is an n-fold generalization, a sample of some of the deep results obtained by Alan Baker in a sequence of papers in the late 1960's (for which he got the Fields prize). If $\alpha_1, \ldots, \alpha_n$ are algebraic numbers different from both 0 and 1, and if $\beta_1, \ldots, \beta_n$ are algebraic numbers such that $1, \beta_1, \ldots, \beta_n$ are linearly independent over the field of rational numbers, then $\alpha_1^{\beta_1}\alpha_2^{\beta_2} \cdots \alpha_n^{\beta_n}$ is transcendental.

Reference: EDM 414/D.

Development 3: The continuum hypothesis. In 1900 there was an International Congress of Mathematicians in Paris, and that's where Hilbert presented his list of 23 problems. The first of those problems was the continuum hypothesis, originally formulated by Cantor. The simplest statement of Cantor's continuum hypothesis (there are other, more general, statements of great interest) is that every uncountable set of real numbers is in one-to-one correspondence with the set of all real numbers, or, in Cantor's notation, that there is no cardinal number between $\aleph_0$ and $2^{\aleph_0}$.

Is the continuum hypothesis true? The question has often been likened to a similar one about Euclid's parallel postulate, and the answer has shocked and annoyed a lot of people, just as the Bólyai-Lobachevsky resolution of the status of the parallel postulate shocked and annoyed many of our grandfathers. In both cases there is a more or less pleasant axiom system (in the present case it is the Zermelo-Fraenkel system of axioms for set theory) and a less pleasant, more complicated, non-obvious additional axiom. If the extra axiom is a consequence of the basic ones, it is true, and all is well; if its negation is a consequence of the basic ones, it is false, and, for better or for worse, the question is definitively answered. The answer, awaited for a long time, turns out to be a subtle and profound intellectual achievement. Gödel proved in 1940 that the continuum hypothesis is not false—it is consistent with the other set-theoretic axioms—and Paul Cohen proved in 1964 that it is not true—it is independent of the other axioms, or, in other words, its negation is also consistent.

Both Gödel and Cohen argue by the construction of a suitable model, but use very different techniques. Gödel starts with a universe of sets that satisfies the Zermelo-Fraenkel axioms and shows that there exists a subuniverse that also satisfies them, and in which, moreover, the continuum hypothesis is true. Cohen's argument is similar but harder. It is reminiscent of Felix Klein's construction of a Lobachevskian plane that endows a Euclidean disk with a new metric. Cohen, like Gödel, starts with a model of set theory and then enlarges it, adjoins new objects to it, in such a way as to "force" the continuum hypothesis to be false. ("Forcing" has become an important technical word in the subject.)

Where does that leave the continuum hypothesis? Many people believe, as Gödel did, that despite the independence of the continuum hypothesis it is in some legitimate sense either true or false—that humanity has not yet thought of the right way to describe the full truth about set theory, and when the full truth comes to be known, when the appropriate additional axioms are found and adjoined to

the present ones, then the continuum hypothesis will become either provable or disprovable. Both schools of thought exist, and you are free to join whichever one you find more attractive. Would it influence your decision if you knew what Gödel believed? He thought it would become disprovable.

Reference: EDM 35/D.

Development 4: Lie groups. Hilbert's fifth problem asked whether some relatively mild assumptions about topological groups were enough to imply a strong conclusion. A topological group is a set that is both a group and a topological space in which the two structures are compatible in the sense that the group operations (multiplication and inversion) are continuous.

A typical example is the set of all 2×2 matrices of the form

$$\begin{bmatrix} x & y \\ 0 & 1 \end{bmatrix}$$

with $x > 0$; the topological structure is that of the right half plane (all (x, y) with $x > 0$), and the multiplicative structure is the usual one associated with matrices. This example has an important special property: it is "locally Euclidean" in the sense that every point has a neighborhood that is homeomorphic to an open ball in 2-dimensional Euclidean space. (Equivalently: every point has a "local coordinate system".) An even more important special property of the example is that the group operations, regarded as functions on the appropriate Euclidean space, are not only continuous but even analytic. That is obvious at a glance; if the matrix

$$\begin{bmatrix} x & y \\ 0 & 1 \end{bmatrix}$$

is identified with the ordered pair (x, y), then

$$(u, v)(x, y) = (ux, uy + v),$$

and

$$(x, y)^{-1} = \left(\frac{1}{x}, -\frac{y}{x}\right).$$

If a group is locally Euclidean, that is, if it can be "coordinatized" at all, then there are many ways of coordinatizing it; if at least one of them is such that the group operations are analytic, the group is called a "Lie group". Hilbert's fifth problem was this: is every locally Euclidean group a Lie group?

The problem is analogous to one in complex function theory. It is relatively elementary that a twice-differentiable function is analytic; the hard theorem is that the conclusion holds under much weaker assumptions. Similarly, it has been known for a long time that if a topological group has sufficiently differentiable coordinates, then it has analytic ones; the Hilbert problem is to draw the same conclusion under much weaker assumptions.

Immediately after the discovery of Haar measure, von Neumann (1933) applied it to prove that the answer to Hilbert's question is yes for compact groups. A little later Pontrjagin (1939) solved the abelian case, and Chevalley (1941) solved the solvable case. The general case was solved in 1952 and 1953 by Gleason, and Yamabe, and, jointly, by Montgomery and Zippin; the answer to Hilbert's question is yes. Gleason offered a new characterization of Lie groups; Montgomery and

Zippin used geometric-topological tools (and Gleason's result) to reach the desired conclusion.

Reference: EDM 406/N.

Development 5: Simple groups. Every group has two obvious normal subgroups, namely the group itself and, at the other extreme, the subgroup that consists of the identity element only. A group is called simple if these are the only normal subgroups it has.

Simple groups are like prime numbers in two ways: they have no proper parts, and every finite group can be constructed out of them. Consider, indeed, an arbitrary finite group, and look at a maximal normal subgroup of it, that is, a normal subgroup that is not included in any other proper normal subgroup. If the group is simple, then that maximal normal subgroup is just the identity, but in any event, no matter what that subgroup is, its maximality implies that the quotient group obtained by dividing out by it is simple. The relation among group, normal subgroup, and quotient is sometimes described by saying that the original group is an extension of the quotient group by the subgroup. In this language, every finite group (except the trivial group, the identity) is an extension of a simple group by a group of strictly smaller order. That statement is the group-theoretic analogue of the number-theoretic one that says that every positive integer (except 1) is the product of a prime by a strictly smaller positive integer.

If the maximal normal subgroup is not trivial, then the procedure just used can be used again; the result is a maximal normal subgroup of the maximal normal subgroup such that the original subgroup is an extension of the second quotient by the second subgroup. The procedure can be repeated so long as it produces non-trivial subgroups; the end product is a decreasing chain (a composition series) of subgroups of the original group with the property that each quotient group obtained by dividing a term of the chain by its successor is simple. A great part of the problem of getting to know all finite groups reduces in this way to the determination of all finite simple groups.

The abelian ones among the finite simple groups are easy to determine—it's a classroom exercise to show that they are just the cyclic groups of prime order. That's the only easy part of this subject. What's hard is to find all non-abelian simple groups. Some examples of simple groups are not hard to come by; among permutation groups, for instance, the most famous ones are the alternating groups of degree 5 or more. For a long time the known simple groups did not exhibit any pattern, and even the simplest questions about them resisted attack. Burnside, for instance, conjectured (in 1911) that every finite non-abelian simple group has even order, and that conjecture stood as an open problem for more than 50 years.

In a spectacular display of group-theoretic power, Feit and Thompson settled Burnside's conjecture (in 1963)—it is true. The proof occupies an entire issue (over 250 pages) of the *Pacific Journal*. It is technical group theory and character theory. Some reductions in it have been made since it appeared, but no short or easy proof has been discovered. The result has many consequences, and the methods also have been used to attack many other problems in the theory of finite groups; a subject that was once pronounced dead by many has shown itself capable of a vigorous new life. Thus, for example, Burnside's dream is by now completely realized: by a gigantic collaborative effort of many mathematicians all over the planet all finite simple groups have been found and can be explicitly described.

Reference: EDM 160/D.

Development 6: The Atiyah-Singer index theorem. How can a linear transformation on a finite-dimensional vector space fail to be invertible? There are two obvious answers: it can fail to be injective (one-to-one) or else it can fail to be surjective (onto). The first kind of failure is the existence of a nontrivial kernel, and the second kind is the existence of a nontrivial cokernel. The kernel of a transformation T is, of course, the null-space, the inverse image of 0; the cokernel is, roughly, the complement of the range, or, precisely, the quotient of the entire space modulo the range. It is an elementary fact of linear algebra that the two kinds of failure always occur simultaneously, and that, in fact, the numerical measures of the two kinds of failure, that is, the dimensions of the kernel and the cokernel, are always equal.

For infinite-dimensional spaces the story is different. If, for instance, T is the unilateral shift on the infinite-dimensional sequence space l^2, the transformation defined by

$$T\{\xi_1, \xi_2, \xi_3, \dots\} = \{0, \xi_1, \xi_2, \xi_3, \dots\},$$

then $\ker T$ is the trivial subspace $\{0\}$, but $\operatorname{ran} T$ is the subspace of all vectors whose first coordinate is 0, and consequently $\dim \ker T = 0$ and $\dim \operatorname{coker} T = 1$.

The *index* of any linear transformation T (including the one just described) is defined by

$$\operatorname{index} T = \dim \ker T - \dim \operatorname{coker} T,$$

whenever that makes sense (that is, whenever it is not equal to $\infty - \infty$). For linear transformations on finite-dimensional spaces the index is always 0, but on infinite-dimensional spaces the index (being a measure of the extent to which a transformation is irreparably non-invertible) can give interesting information.

The Atiyah-Singer index theorem has to do with the "analytic" index just defined. The first instance of it that most of us run into is the Cauchy integral formula thought of as the computation of a winding number. Another instance is the Riemann-Roch theorem. Compact Riemann surfaces such as the sphere and the torus arise in complex function theory, and the Riemann-Roch theorem (which goes back to a paper by Roch in 1865) is about the dimensions of certain vector spaces of meromorphic functions on such Riemann surfaces; its conclusion is a formula for that dimension. The general Atiyah-Singer theorem is a generalization of the Riemann-Roch theorem. It deals with smooth compact manifolds more general than Riemann surfaces. Elliptic differential operators acting on smooth functions defined on such manifolds have two numbers associated with them. One of them is the analytic index, as above, and the other is the topological index, which is something else. The topological index is related to K-theory, and, in particular, in more classical contexts, to the Euler characteristic. The achievement of the Atiyah-Singer theorem (1963) is that those two indices have the same value —that, in other words, a property of great analytic importance, with a purely analytic definition, is in fact almost completely determined by the topological properties of the underlying manifold.

The equation just referred to is only a small part of the joint work of Atiyah and Singer. The results of that work are the deepest and broadest, and, for me as a reporter, much the hardest of the contents of this report; they are not just a theorem but a theory, a context, a point of view that enters, influences, and is influenced by many parts of mathematics. Writing about the spectacular successes of research in differential geometry during the last fifty years, Osserman called the

Atiyah-Singer index theorem "a grand synthesis of analysis, topology, and geometry leading, in particular, to a new way of viewing the Gauss-Bonnet theorem: not as an isolated result, but as one instance of a larger scheme of things." That synthesis was probably a major factor in Michael Atiyah's being knighted in England and Isadore Singer's being awarded a presidential medal in this country.

Reference: EDM 236/H.

Development 7: Fourier series. It is a historical misfortune (which was responsible for almost 200 years of barking up the wrong tree) that Fourier series were discovered before convergence. Fourier series are a vital part of both classical and modern analysis; they are important for both abstract theory and concrete applications. They arise in topological groups and in operator theory; they have their origins in problems about vibrating strings and about heat conduction.

In their most classical manifestation, Fourier series have to do with numerical-valued functions on the line (it is usually best to let them be complex-valued) that are periodic of period 2π and integrable on the interval $[0, 2\pi]$. The Fourier series of such a function is the infinite linear combination of the exponential functions e^{inx}, $n = 0, \pm 1, \pm 2, \ldots$, with coefficients determined by integrating the given function against them. (Alternatively the Fourier series uses sines and cosines running in one direction only; the complex version, however, is algebraically simpler to manipulate.)

Trigonometric polynomials (in either real or complex form) are familiar objects and are computationally accessible; surely nothing but good could come out of representing more difficult functions as limits of such polynomials. It seemed natural, therefore, to hope that the "sum" of the Fourier series associated with a function f would be "equal" to f, and, in any event, to ask for which functions that would occur. The answer, it was hoped, was that good functions would have good series, and the history of the subject has been strongly influenced by that hope.

When limits began to be understood, "sum" and "equal" were interpreted in the sense of pointwise convergence; the more fruitful and usable concepts of weak convergence and convergence with respect to a norm came along only after the mathematical community was irretrievably committed to research in the pointwise direction.

How good does a good function have to be? Differentiability is good enough, but, it turns out, continuity is not; there are continuous functions whose Fourier series diverges at a point (du Bois Reymond, 1876), and, in fact, at many points. Must the Fourier series of a continuous function converge almost everywhere? That was an unsolved problem for many years.

Kolmogorov showed that if all that is assumed is that f is integrable on $[0, 2\pi]$, then it could happen that the Fourier series of f diverges almost everywhere (1923), or even everywhere (1926). The biggest question along these lines was asked by Luzin, and it remained unanswered for 50 years: if f is square integrable on $[0, 2\pi]$, does it follow that the Fourier series of f converges to f almost everywhere? Repeated failure to prove the affirmative led to the official state religion among the cognoscenti in the 1950s and 1960s: the answer must be no.

The answer is yes. The first proof is due to Carleson (1966). A remarkable feature of Carleson's achievement is that it uses no unknown techniques; it just uses the known ones better. It depends on an ingenious push-me-pull-you way of

selecting subintervals. It is as if Carleson had power enough to replace everyone else's ε by ε^2, and that did the trick.

Reference: EDM 167/H.

Development 8: Diophantine equations. Hilbert's tenth problem concerned the solvability of Diophantine equations. The problem was to design an algorithm, a computational procedure, for determining whether an arbitrarily prescribed polynomial equation with (positive) integer coefficients has (positive) integer solutions. (The modifier "positive" is technically convenient and does no harm.)

What does it mean to say there is an algorithm for deciding solvability? A reasonable way to answer the question is to offer a definition of computability for sets and functions, and then to define an algorithm in terms of computability. The concept of computability has received a lot of attention; it has several different but logically equivalent definitions all in accord with the intuitive notion that the word suggests.

Suppose now that $\{E_1, E_2, E_3, \dots\}$ is an effective enumeration of all the polynomial equations under consideration, and let S be the set of those indices k for which E_k has a solution. Hilbert's problem (is there an algorithm?) can be expressed by asking whether S is a computable set. The answer is no. The answer was a long time coming: it is the result of the cumulative efforts of J. Robinson (1952), M. Davis (1953), H. Putnam (1961), and Y. Matijasevič (1970).

The central concept in the proof is that of a Diophantine set, and the major step proves that every computable set is Diophantine. The techniques make ingenious use of elementary number theory (e.g., the Chinese remainder theorem, and a part of the theory of Fibonacci numbers). The proof exhibits some interesting Diophantine sets whose Diophantine character is not at all obvious (e.g., the powers of 2, the factorials, and the primes).

One way to prove that S (the index set of the solvable equations) is not computable is by contradiction. If S were computable, then it would follow (by a slight bit of additional argument) that each particular Diophantine set (i.e., the solution set of each particular equation) is computable, and hence (by the "major step" mentioned above) that the complement of every Diophantine set is Diophantine. The contradiction is derived by exhibiting a Diophantine set whose complement is not Diophantine.

The last step uses a version of the familiar Cantor diagonal argument. The idea is "effectively" to enumerate all Diophantine sets, as $\{D_1, D_2, D_3, \dots\}$ say, to prove that the set $D^* = \{n: n \in D_n\}$ is Diophantine (that takes some argument), and, finally, to prove that the complement of D^* is not Diophantine—and that's where Cantor comes in.

Reference: EDM 100.

Development 9: Banach bases. In calculus (beginning in the 17th century) we learn how to find maxima and minima of numerical functions defined on familiar domains such as intervals in the line and rectangles in the plane. In a later subject called the calculus of variations (beginning in the 18th century) we try to find maxima and minima of numerical functions whose domains consist of sets of functions. Most famous example: for each path joining two points in space there is a definite time that it takes for a particle to slide along the path from the first point to the second—what is the minimum of all those times? (The minimum time

is attained for the celebrated path called the brachistochrone—*brachisto* for shortest and *chrone* for time.) Problems such as that were among the ones that gave rise to functional analysis, the part of analysis in which the domains of the functions studied are sets of functions, subsets of function spaces, and in which algebraic and topological methods are freely used.

Another point of view on functional analysis is that it is an infinite-dimensional generalization of linear algebra. The first idea was to replace vectors and the additions involved in applying matrices to them by functions and the integrations involved in applying kernels to them. The first systematization of the subject (around the 1920's) was proposed by Banach and Wiener (independently)—the result was the theory of Banach spaces, and its many wildly generalized outgrowths. The study of Banach spaces is a typical instance of the axiomatic method in mathematics—abstract and general, with roots in the concrete and special. It is called too general by some and not general enough by others. In any event it is still a living part of mathematics, and major progress was made in it relatively recently.

One of the earliest questions about Banach spaces was the basis problem, raised by Banach himself in his book (1932). A sequence of elements in a Banach space is a Schauder basis for the space in case every vector has a unique expression as a convergent infinite linear combination of the terms of the sequence. The countability built into the definition of the word "sequence" implies that if a Banach space has a basis, then it is separable (that is, has a countable dense set). The basis problem, which was outstanding for 40 years, was the converse: does every separable Banach space have a basis? Each space that ever came up in analysis had one, and yet a proof that that had to be so remained elusive.

A classically important concept in the study of Banach spaces is that of a compact (completely continuous) operator, that is, a linear transformation between Banach spaces with the property that the closure of the image of the unit ball is compact. The easy compact operators are the ones of finite rank (that is, the ones for which the range is finite-dimensional); the next easiest ones are the (uniform) limits of operators of finite rank. If a Banach space is "good", then every compact operator mapping into it is such a limit (in technical language: the space has the approximation property), and, in particular, if a Banach space has a basis, then it has the approximation property.

The basis problem was solved by Per Enflo in 1973. The solution turned out to be negative: there exists a separable Banach space that does not have the approximation property. The technique is constructive; it is a combinatorial way of constructing and putting together infinitely many finite-dimensional Banach spaces.

Reference: EDM 39/A.

Development 10: Manifolds. A 2-manifold is a topological space that is locally like 2-dimensional Euclidean space (that is, every point has a neighborhood homeomorphic to an open disk in the plane) with the properties that the locally Euclidean patches are glued together continuously and that there are not too many of them (meaning that the entire space is separable). The definitions of 3-manifolds, 4-manifolds, 5-manifolds, etc. are exactly the same—just replace 2 by 3, 4, 5, etc., (and replace "plane" by 3-space, 4-space, 5-space, etc.).

A 2-manifold can be large (it can, for instance, be the entire plane), and even if it doesn't look large it may be intrinsically large in the sense that many paths in it converge to points that don't exist in it—as is the case with an open disk. The manifolds whose study is more promising are the compact ones.

Can all possible compact 2-manifolds be listed? Yes, and the simplest ones to list are the so-called orientable ones. One of them is simply connected (the sphere), another has genus 1 (the torus), another has genus 2 (a doughnut with two holes), etc.—and all possible orientable 2-manifolds are like one of these (in the sense of homeomorphism). The classification problem for compact 2-manifolds is classical and was solved long ago. (The 1-dimensional case can safely be left as homework.)

For higher dimensions things are tougher to see and tougher to do. Surprisingly, however, when the dimension is 5 or greater, much of the answer is known—simply connected 5-manifolds are understood and classified by homotopy theory. For 3-manifolds the problem is unsolved—that's what the celebrated Poincaré hypothesis is about. For 4-manifolds the story is different. Two great victories were won by Michael Freedman and Simon Donaldson in 1982 (and each was awarded a Fields medal at the Berkeley Congress in 1986).

It turns out that every oriented 4-manifold is associated with a symmetric integer matrix ("the intersection matrix") of determinant ± 1, and Freedman showed that all such matrices can indeed occur. The correspondence between matrices and manifolds is one-to-one half the time and two-to-one the rest of the time. The final result is a complete classification of all simply connected oriented 4-manifolds. If we think of a basketball (the surface of a sphere) as a 2-dimensional object (never mind that we usually see it embedded in 3-dimensional space), Freedman's work gives us as much insight about 4-dimensional basketballs as we have about the honest-to-goodness ones.

Donaldson, on the other hand, showed that if a simply connected oriented 4-manifold has a smooth differential structure and if its matrix is positive definite, then the matrix must be equivalent to the identity matrix. That's a very strong conclusion; it shows that the theories of topological and differential 4-manifolds are radically different.

Presenting Freedman's work, John Milnor said that the proofs of the results are extremely difficult. Presenting Donaldson's work, Michael Atiyah said that it opened up an entirely new area, and he went on to say that Donaldson's youth and mathematical power are "an indication that mathematics has not lost its unity, or its vitality".

Reference: MI 5/3/39.

Development 11: The Bieberbach conjecture. Some problems are interesting not because they are interesting but because they are elusive. The Fermat question is probably the best case in point. Nobody really wants to know the mere answer to that one—what mathematicians want to know is why they don't know the answer. The Bieberbach conjecture about schlicht functions was, in context, perfectly natural to the specialist—to most outsiders, however, it seemed like a curious technicality whose main claim to fame was that it remained unanswered.

What it concerned was the class of all those functions that are given by power series of the form

$$z + \sum_{n=2}^{\infty} a_n z^n$$

and that are schlicht (meaning analytic and injective or, in the more classical word, univalent) in the open unit disk. The class constitutes a "normal family" (a

compactness property), which implies that, for each n, the coefficient a_n remains bounded throughout the class. Bieberbach, studying the subject in 1916, proved that in fact $|a_2| \leqq 2$ for every function in the class. One particular function in the class, the Koebe extremal function defined by $a_n = n$, shows that the upper bound 2 is best possible (and is, in fact, attained). The Bieberbach conjecture was that, more generally, $|a_n| \leqq n$ for all n.

Löwner proved in 1923 that the conjecture is true for $n = 3$; in 1955 Garabedian and Schiffer proved it for $n = 4$; in 1968 Pedersen and Ozawa proved it for $n = 6$; in 1972 Garabedian, Pedersen, and Schiffer proved it for $n = 5$; and in 1973 Ozawa and Kubota proved it for $n = 8$. Progress was slow and not promising.

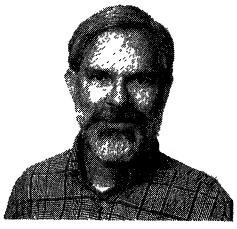
The breakthrough that transformed the conjecture into a theorem came in 1984 when Louis de Branges offered a proof of the general case, a proof of several hundred pages, based on his theory of square summable power series. His original proof contained minor errors, but they were correctable and soon corrected. The experts were still a little bothered, but their discomfort didn't last too long. What they didn't like was the reliance of the proof on the seemingly irrelevant methods of functional analysis, and, sure enough, ultimately the Leningrad seminar on geometric function theory produced a proof that Bieberbach would have liked. The proof is shorter and more perspicuous than its special predecessors for the cases $n = 5$ and $n = 6$. Square summable power series are gone, but the main structural insights of de Branges are still the ones that make things go. We don't, glory be to glory, lose them all.

Reference: MI 8/1/40.

Epilogue. The answer to the question in the title is clearly and decisively no.

Two Theorems, Two Sculptures, Two Posters

HELAMAN FERGUSON, *Supercomputing Research Center, Bowie, Maryland*



HELAMAN ROLFE PRATT FERGUSON has an A.B. in liberal arts from Hamilton College and a Ph.D. in mathematics from the University of Washington. He is a full Professor of Mathematics at Brigham Young University, currently with the Supercomputing Research Center, Bowie, Maryland. His published mathematical work is primarily in applying matrix groups to algorithm design. Helaman learned to work with stone as a young apprentice, studied painting as an undergraduate and had graduate courses in sculpture. His sculpture studio is now in Maryland; his collected sculptural work is direct carving of stone and bronze. Helaman is in the 1988 and 1989 Guinness Book of World Records for juggling a distance of fifty miles. Helaman and his wife Claire are raising six sons and one daughter. Claire is an artist; she understands mathematicians very well and has observed, "I know what you guys are doing. First you shoot the arrow and then you draw the bullseye."

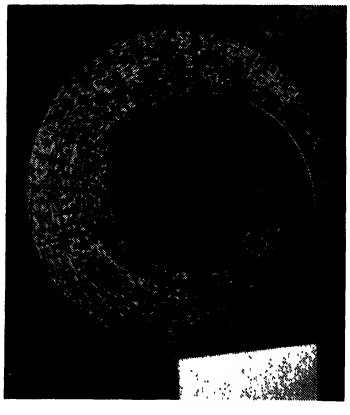


FIG. 1. *Umbilic Torus NC* Front View; the closed curve of cusps goes round the torus thrice the long way and once the short way.

Art / Science. Mathematics is both an art form and a science. This duality integrates our social reflexes to differentiate disciplines and people. Mathematics is a language of physics and a tool of engineering. In manufacturing engineering there is something called a critical part. This part is so crucial to the operation of the whole that its failure would be a disaster. Such parts are made with the highest quality of attention and materials, and a complete pedigree is maintained of the source, handling, and steps of manufacture. There are crucial mathematical theorems that are worth this high level of attention and celebration.

I think that the great myths of our century are mathematical equations; the heroes of those myths are equals signs and approximations. Haven't we all flown comfortably upon the implications of Navier-Stokes equations? We consistently build effective machinery by testing mathematical models or 'ghost machines' in computer simulations. We have impressive engineering professions motivated by function. Function and efficiency were never everything. But mathematics has

another side, motivated by a successful aesthetic quite its own. What if the resource, intelligence, sophistication, and energy of mathematics were developed for aesthetic creation of perfectly useless and priceless artifacts?

Mathematician / Sculptor. I believe it is feasible to communicate mathematics along aesthetic channels to a general audience. Mathematical beauty as well as mathematical truth is distinctive and highly developed. A very rigorous mathematician submits that "in all mathematics there is a mysterious element: the astonishing concurrences and ties between objects and theories which at first glance seem far apart" [Arnold]. Sometimes I have succeeded in capturing the truth, the beauty, and this mysterious element in my sculpture [Cannon, Sakane].

I am only one kind of mathematician, one kind of sculptor. My mathematical processes tend to be constructive and algorithmic, [J. Algorithms paper, Anaheim talk]. An algorithm is a construction involving step by step putting together and organization of resources. Some mathematicians work analytically by breaking things down. On the other hand, my sculpture process tends to be destructive. I begin with raw material such as stone and systematically destroy the parts I don't want to see or feel. In sculpture jargon, this is called direct carving. It is a subtraction process. Some sculptors work mostly by building up. It is helpful to draw these distinctions, but they are not hard and fast.

The first sculpture I tell about is direct carving old style; I direct carved pretty much the same way the sculptor-carvers of Willendorf, Hierakonpolis, Ur, Harappa, Yazilikaya, Imdugud, Warka, Sidon, Chavin, Gandhara, Yaxchilán, Mathura, Vijayanagar, T'ien-hung Shan, T'ang, Souillac, and Naumberg did. I carved much the way Myron, Praxiteles, Sluter, Michelangelo, Bologna, Bernini, Coysevox, Story, Thorvaldsen, Greenough, Rodin, Taft, Brancusi, Hepworth, Moore, Noguchi, or their assistants did. The second sculpture was done by direct carving new style: I direct carved using an interesting combination of mathematics and technology not available to the above masters.

First theorem: the classification of surfaces. The homeomorphism problem of surfaces is posed as follows: give a short list of easily computable criteria which are different for surfaces not homeomorphic and which are the same for surfaces which are homeomorphic. The fundamental theorem of the topology of compact surfaces states that two compact surfaces are homeomorphic if and only if they have the same number of boundaries, handles, and cross-caps. This theorem completely solves the homeomorphism problem for surfaces. Surfaces have dimension $n = 2$. The same problem for higher dimensions $n \geqslant 3$ is still open, [Scott, Thurston].

The *Torus with Cross-Cap* sculpture presents one handle, one cross-cap, and, if one counts the footprint, one boundary. A *Torus with Cross-Cap* contains the germ of the fundamental theorem of the topology of surfaces. When a *Torus with Cross-Cap* is dug out of somebody's rubble in ten thousand years, the knowledge of this fundamental theorem can be celebrated again. Theorems are time invariant, so I like to sculpt in geological time scale materials.

The proof of $3x = x + h$; $2x \neq h$. At first sight $3x = x + h$; $2x \neq h$ looks like an obvious error. The usual algebra of cancelling an x in $3x = x + h$ gives $2x = h$. Algebraic statements are not necessarily context-free. These algebraic statements are taking values in geometric structures; x represents a cross-cap on a sphere,

+ represents a connected sum, h represents a handle on a sphere, and = represents a reversible deformation. Other things about this algebra, such as cancellation, are unusual. Three cross-caps on a sphere, $3x$, is a torus with a cross-cap, $x + h$, but a Klein bottle, $2x$, is not a torus h.

Follow the visualization indicated by the accompanying drawings; explain your visualization to someone else. Begin with a sphere, cut out boundaries, sew on handles and sew on cross-caps.

Here are two kinds of pictures: sew-ups (leading to algebra) and curve drawings (leading to sculpture). A sew-up is a planar region with marked boundaries and directions for sewing the boundaries together. For convenience, the directions will rely upon personal experience with physical paper and cloth, but only up to a point. A new, non-physical ingredient is the possibility of self-intersection. The curve drawings are shorthand symbols for a thing visualized in the mind and body somehow; a thing some might like to think has independent existence, but a thing that takes time and work to visualize. If a visualization doesn't develop one day it will the next. Distinguish when a curve outlines a surface tangent to lines of sight, describes a surface self-intersection, lies under part of the surface, or informs a cross-section.

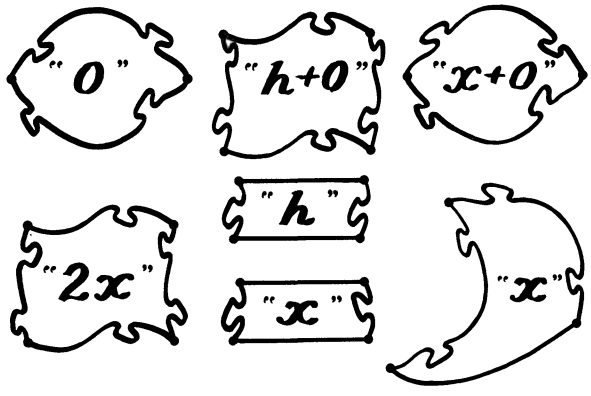


FIG. 2. Six Sew-ups: A Sphere, 0; A Torus, $h + 0$; A Projective Plane, $x + 0$; A Klein Bottle, $2x$; A Handle, h; A Möbius Band, x; A Cross-Cap, x. Sew matching sides of the self inter-locking puzzles.

Recognize that the algebraic expressions written on the drawings are consistent with the definition of equality given by reversible deformations and the definition of addition given by the connected sum. Reversible deformation means stretching and shrinking, neither cutting nor tearing, but self-intersections are allowed. The connected sum refers to the process of joining two surfaces along the boundary of a removed disc. There are four interesting reversible deformations and each has a sequence of curve drawings. Each curve drawing in the sequence presents a new idea in the reversible deformation. The equality between the beginning and ending structures may surprise at first but after practice the deformation will seem obvious.

Between a Möbius band and a cross-cap, stretch the boundary of the Möbius band as drawn up on the lower part and down on the upper part. The self-intersection line of the cross-cap appears naturally when this is done. The cross-cap is a

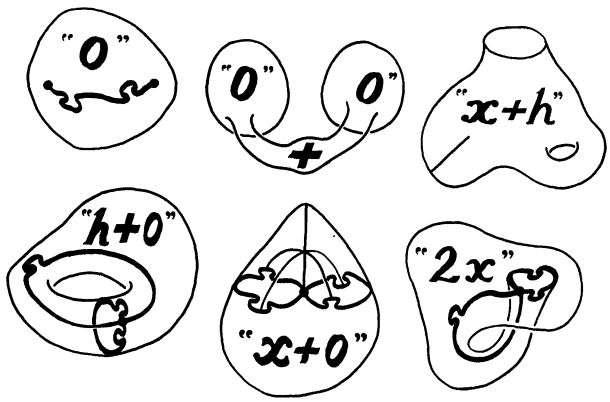


FIG. 3. Six Curve Drawings: A Sphere, 0; A Connected Sum of Two Spheres, $0 + 0$; A Sphere with a Boundary, a Handle, and a Cross-Cap, $x + h$; A Torus, $h + 0$; A Projective Plane, $x + 0$; A Klein Bottle, $2x$.

projective plane with a disc removed, the only difference between a Möbius band and a cross-cap is the size of the disc hole and where it is.

In the catalytic presence of a cross-cap drawing, think of one end of the handle of the torus as a sort of vacuum hose snorkel elephant trunk tornado thing that moves its snout as a hole around in the surface. When it moves across the cross-cap it develops a self-intersection but retains its hole to reveal a Klein bottle structure.



FIG. 4. Six More Curve Drawings: A Möbius Band, x; A Cross-Cap, x; A Sphere with Two Cross-Caps, $2x$; A Sphere with Three Cross-Caps, $3x$; A Torus with a Cross-Cap, $x + h$; A Klein Bottle with a Cross-Cap, $2x + x$. The last three are the same.

The AMS Sculpture: *Torus with Cross-Cap and Vector Field*, Direct Carving $x + h$. In November 1986 I acquired a couple of tons of Carrara white marble and carved what eventually became the present AMS *Torus with Cross-Cap and Vector Field* from one of the 1200 lb chunks. A stone carver would call this spec work. It

was not specifically commissioned and had no immediate home; financially (as usual), I was going out on a rock. I was motivated by the joy of the thing, the same reason I do mathematics. If the finances work out that is great, because then I can continue.

Carrara marble comes from the coastal mountains of northwest Italy south of Genoa, north of Livorno. More specifically, southeast of La Spezia, north of Lucca, east of the towns of Carrara and Pietrasanta, are mountains bearing a close grained, generally white stone of consistently high quality. Many generations of families have been quarrying this stone for building purposes; the effects of their labors, the exposed white disemboweled mountains, can be seen from far at sea and space, as satellite photographs reveal. (These mountains would be a good place for a really large torus with cross-cap.) Fine art sculpture does not support the stone industry in Italy or anywhere else; occasional good blocks become available for reasonable prices with the right connections and for unreasonable prices without.

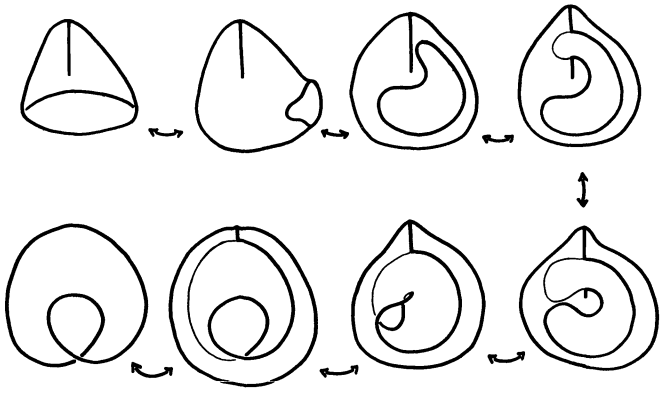


FIG. 5. A Möbius Band is a Cross-Cap, $x = x$. The cross-cap self-intersection line appears naturally when the boundary is crimped toward the center of gravity.

I carved *Torus with Cross-Cap and Vector Field* in eleven days, working 10–16 hours a day. Stone carving is hard work even with modern electric and air tools; it firms muscle and tests courage. Carrara marble is beautiful to carve. It flakes rather than breaks or chunks; it shows a pretty pressure wave as the pneumatic tool penetrates.

I will describe a few details of the process of direct carving $x + h$. The first stage was to get the idea for the work somewhere in mind and body. For me, that began twenty six years earlier when as an undergraduate at Hamilton College I saw an elegant paper by another undergraduate at Caltech [Crawley]. Those topological ideas percolated in my system for years. They were there when I was behaving like an artist and they were there when I was behaving like a scientist. Long term understanding of these ideas became an important source of the myriad decisions in my final conversation with the stone.

The second stage was having the opportunity, which raw materials, tools, and experience afford. But these stages pale temporarily when faced with the problem

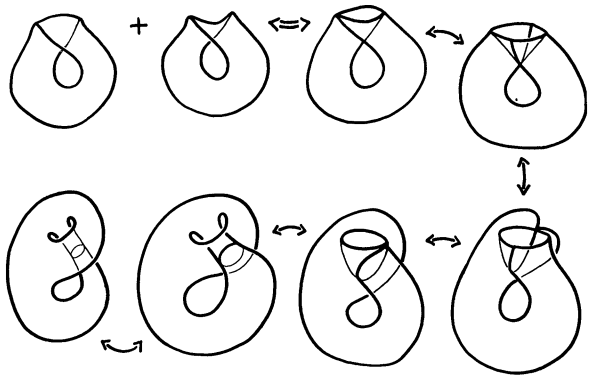


FIG. 6. Sew two Möbius Bands, $x + x$, along their boundaries to get a Klein Bottle, $2x$. The next figure has further detail.

of physically moving a 1200 lb stone by oneself. Remember the integral over distance definition of work with the little differential ds? It is interesting to move a big stone a long way, a very little at a time, with just one crowbar and much patience. This primitive move was necessary to reposition the stone to access it with a four wheeled frame hoist.

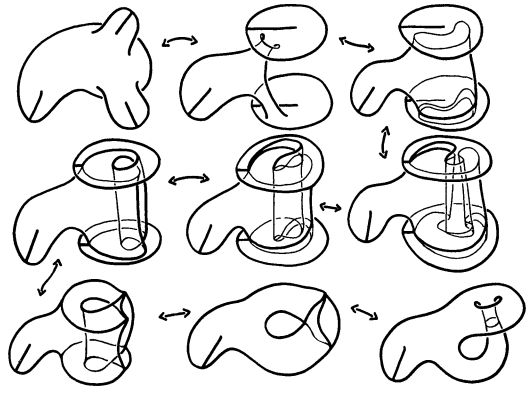


FIG. 7. Between three Cross-Caps on a Sphere and a Cross-Cap on a Klein Bottle, $3x = x + 2x$.

The third stage was positioning the stone, blocking it up in various ways to get a good look at it, and then carving the hole. A practical reason for doing this first was to check out the integrity of the stone and reveal any flaws before making a big investment of time and energy. At the same time I made inroads on the cross-cap as well for the same reason. There was a natural spall on the side of the stone where I had decided to uncover the cross-cap. All carving to this point was done with a hand-held pneumatic hammer with a sleeve insert for taking a traditional chisel, usually a point, and a single carbide tipped hammer drill for the guide hole. I had satisfied myself on the integrity of the stone; no fissures appeared along which the stone could mysteriously and silently separate into two or more pieces.

The block was somewhat rectangular, so I used my hammer drill again, drilling a series of holes comprising the general outline of the piece, face on, as it appears in the AMS poster. I tapped soft iron into the drill holes, uniformly all the way around, to exert pressure on the hole weakened stone and shear off the unwanted corners. A similar smaller scale operation involved making parallel carbide saw cuts about an inch deep and an inch or less apart, then popping the blocklets off with a parting tool.

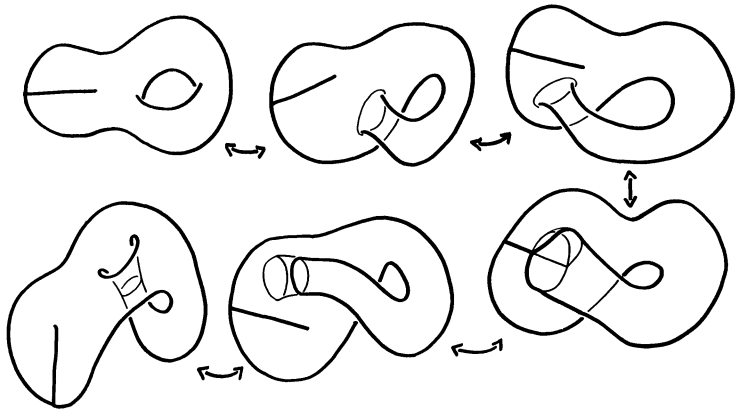


FIG. 8. Between a Cross-Cap on a Klein Bottle and a Cross-Cap on a Torus, $x + 2x = x + h$. The Cross-Cap as Catalyst.

After that initial gross chamfer of the block I proceeded with a pneumatic hammer and a point. I was shaping a surface that had a local 2-frame almost everywhere. I didn't require the 2-frame to be orthogonal and in the case of the crystal structure of this stone more like sixty degrees was better. So my cutting pattern was to follow two adjacent sides of a parallelogram, first stroke one side and then the other side. This operation popped off a parallelopiped shaped flake, quantitatively over six hundred pounds of them. The resulting carving was overall within a half inch or so of the final and had an interesting pineapple look and texture. I worked sensitively especially in the region of the self-intersection line of the cross-cap; punching through or splitting would ruin the entire piece. I wanted to approximate the line of intersection as closely as possible and reveal the expressive translucence of the white marble in that narrow region.

The next stage was grinding with a set of silicon carbide wheels and diamond discs. Dust control is a serious problem at this stage. Marble is mostly the relatively harmless mineral calcium carbonate. It is unhealthy but not usually lethal to inhale some marble dust. Granite and other silicate based stone dust circulating in the lungs can over the long term cause cancerous lesions. I used a pair of snorkels to draw the dust laden air into a filter system which captured the dust and vibrated it periodically into a barrel.

Both of these stages, flaking and grinding, were performed gradually over the entire piece. This meant not penetrating too deeply in any particular area but working with the whole simultaneously. Balance was maintained throughout.

Repositioning and reblocking of the piece was required to access the different faces; this was time consuming even though the piece was gradually losing weight.

I brought the final surface up by sanding with successively finer grits, 60 mesh, 120 mesh, 320 mesh, and finally about 400 mesh. I used flap wheels and discs mounted on one of my one-third horsepower flexible shaft tools at 2800 rpm; silicon carbide, zirconium oxide, and diamond were the abrasive materials. The surface was close to an egg shell polish; a very smooth surface which I intended to texture for mathematical as well as tactile reasons.

Mathematically there is an interesting difference between a cross-cap self-intersection of two surfaces and two abutting corners. The latter had already shown up in sculpture and could be had with two straight saw cuts and no particular care. To distinguish this cross-cap here I articulated the continuation of one surface through the other with a corresponding horizontal vector field on the one and a vertical vector field through the other. I developed the vector field theme by cutting many parallelish polished grooves over the entire surface. To do this I used a tungsten carbide conical ball end cutter rotating at 27,000 rpm. These strokes cover the entire surface of the *Torus with Cross-Cap* and now *Vector Field*. I also used this high speed tool to refine the approximation to the cross-cap as much as I dared. One slip and ruin. The final cross-cap has a lovely translucent effect when backlit.

This summarizes the direct carving process. I emphasize that concentration is essential throughout the whole process. The concentration aspect is very much like doing a long and complex computation with pencil and paper.

The AMS Poster. The American Mathematical Society (AMS) is a professional society primarily devoted to the publication of new mathematical research. The AMS has over 25,000 members, publishes six journals and fifteen book series. It was founded in 1888 and in the century since has contributed to what has been a golden age of mathematics. For the occasion of the 1988 Centennial the AMS organized a special summer meeting at its headquarters in Providence, Rhode

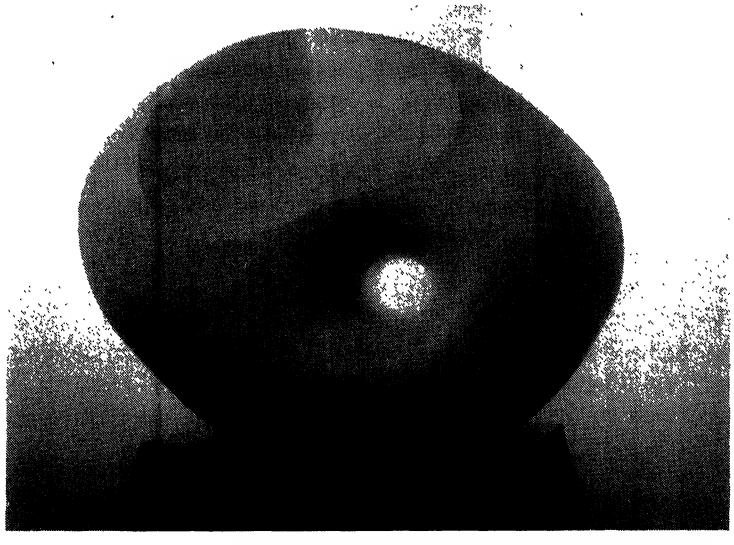


FIG. 9. *Torus with Cross-Cap and Vector Field*, AMS Building, Providence, Rhode Island.

Island [Jackson]. The Mathematical Association of America (MAA) is a professional society with more generalized and public goals than the AMS. The MAA has over 32,000 members and publishes three journals, a newsletter, and seven book series. These two societies hold their meetings jointly and enjoy a spirit of complementary cooperation. The 1987 Summer Meeting of the AMS-MAA was at the University of Utah, only fifty miles north of my recently completed five hundred and fifty pound *Torus with Cross-cap and Vector Field*. This gave MAA President Len Gillman and others opportunity to see the work and approve the purchase. The AMS accepted this MAA gift in the words of AMS President Dan Mostow, "A feast for the eyes as well as the mind."

Many universities and corporations have framed and hung the elegant poster that the AMS published in 1988 featuring this white marble carving on a royal blue background [Pitcher].

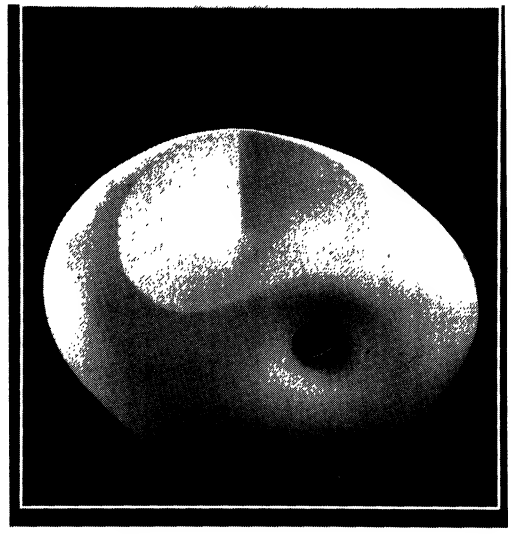


FIG. 10. AMS Centennial Poster, *Torus with Cross-Cap and Vector Field*.

Second theorem: Orbits of $GL(2, \mathbf{R})$ on binary forms. The set of real $n \times n$ matrices with non-zero determinant is a group with the usual matrix multiplication, associativity, identity, and inversion. Call this group $GL(n, \mathbf{R})$. The case of $n = 1$ is the multiplicative group of non-zero real numbers. The case of $n = 2$ has richer structure; $GL(2, \mathbf{R})$ is a very typical and very important group. We will not consider $n \geqslant 3$ here.

$GL(2, \mathbf{R})$ consists of exactly the real invertible 2×2 matrices:

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \qquad \alpha\delta - \beta\gamma \neq 0.$$

These matrices act on column vectors by left matrix multiplication to give the

substitution

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \alpha x + \beta y \\ \gamma x + \delta y \end{pmatrix}.$$

This substitution into quadratic and cubic binary forms gives representations of 2×2 matrices as 3×3 matrices and 4×4 matrices respectively. The quadratic case represents $GL(2, \mathbf{R})$ as a subgroup of $GL(3, \mathbf{R})$; call this group action $GL(2, \mathbf{R}) \times \mathbf{R}^3$. The cubic case represents $GL(2, \mathbf{R})$ as a subgroup of $GL(4, \mathbf{R})$; call this group action $GL(2, \mathbf{R}) \times \mathbf{R}^4$. These matrix subgroups act by right multiplication on the coefficient row vectors of the corresponding binary forms.

The quadratic and cubic binary forms and their coefficient row vectors are,

$$ax^2 + bxy + cy^2 \leftrightarrow (a, b, c) \in \mathbf{R}^3$$

and

$$ax^3 + bx^2y + cxy^2 + dy^3 \leftrightarrow (a, b, c, d) \in \mathbf{R}^4.$$

The word binary refers to the two variables x, y. Substitute

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \alpha x + \beta y \\ \gamma x + \delta y \end{pmatrix}$$

into each form, expand, and collect $x^\lambda y^\mu$-monomial terms,

$$\begin{aligned} ax^2 + bxy + cy^2 &\to (a\alpha^2 + \alpha b\gamma + c\gamma^2)x^2 + (2a\alpha\beta + b(\alpha\delta + \beta\gamma) + 2c\delta\gamma)xy \\ &\quad + (a\beta^2 y^2 + b\beta\delta + c\delta^2)y^2, \\ ax^3 + bx^2y + cxy^2 + dy^3 &\to (a\alpha^3 + \alpha^2 b\gamma + \alpha c\gamma^2 + d\gamma^3)x^3 \\ &+ (3a\alpha^2\beta + \alpha^2 b\delta + 2\alpha b\beta\gamma + 2\alpha c\delta\gamma + \beta c\gamma^2 + 3d\delta\gamma^2)x^2 y \\ &+ (3a\alpha\beta^2 + 2\alpha b\beta\delta + \alpha c\delta^2 + b\beta^2\gamma + 2\beta c\delta\gamma + 3d\delta^2\gamma)xy^2 \\ &+ (a\beta^3 + b\beta^2\delta + \beta c\delta^2 + d\delta^3)y^3. \end{aligned}$$

These new coefficients are linear in the original coefficients $(a, b, \ldots)$ but non-linear in the $\alpha, \beta, \ldots$, since only powers of the x, y were involved. The 2×2 matrix

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in GL(2, \mathbf{R}),$$

has the representations

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mapsto \begin{pmatrix} \alpha^2 & 2\alpha\beta & \beta^2 \\ \alpha\gamma & \alpha\delta + \beta\gamma & \beta\delta \\ \gamma^2 & 2\delta\gamma & \delta^2 \end{pmatrix} \in GL(3, \mathbf{R}),$$

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \mapsto \begin{pmatrix} \alpha^3 & 3\alpha^2\beta & 3\alpha\beta^2 & \beta^3 \\ \alpha^2\gamma & \alpha^2\delta + 2\alpha\beta\gamma & 2\alpha\beta\delta + \beta^2\gamma & \beta^2\delta \\ \alpha\gamma^2 & 2\alpha\delta\gamma + \beta\gamma^2 & \alpha\delta^2 + 2\beta\delta\gamma & \beta\delta^2 \\ \gamma^3 & 3\delta\gamma^2 & 3\delta^2\gamma & \delta^3 \end{pmatrix} \in GL(4, \mathbf{R}).$$

The corresponding determinants are $(\alpha\delta - \beta\gamma)$, $(\alpha\delta - \beta\gamma)^3$, $(\alpha\delta - \beta\gamma)^6$, so the representations are as invertible as the original matrix.

The $GL(2, \mathbf{R})$ orbit of a form is the set of all $GL(2, \mathbf{R})$ substitutions in the variables x, y of the form. Equivalently, the $GL(2, \mathbf{R})$ orbit is right multiplication of a coefficient row vector by all the representation matrices. An orbit is a subset of the set of all forms, two different orbits are disjoint, the union of all the orbits is a partition of the set of all forms. In the case of the $GL(2, \mathbf{R})$ orbits of quadratic and cubic forms there are only finitely many orbits. For higher degree forms there are infinitely many $GL(2, \mathbf{R})$ orbits.

The classification of the binary forms problem is posed as follows: determine the geometric structure of the $GL(2, \mathbf{R})$ orbits. The classification of binary quadratic forms is a theorem in every calculus book and will be done another way below. The theorem for binary cubic forms is less well known but due to [Cayley, Zeeman] and is the subject of my *Umbilic Torus* sculpture and the discussion below. The same problem for binary forms of degree higher than three is still open [Mumford].

The Cayley-Zeeman construction. A simple question to ask about a polynomial in one variable is, are there any repeated roots?

This question has a wonderful answer: a root is repeated if and only if the coefficients of the polynomial satisfy a certain multivariable polynomial called the discriminant. The quadratic equation

$$ax^2 + bx + c = 0$$

has repeated roots if and only if

$$b^2 - 4ac = 0.$$

The cubic equation

$$ax^3 + bx^2 + cx + d = 0$$

has repeated roots if and only if

$$b^2c^2 - 4ac^3 - 4b^3d + 18abcd - 27a^2d^2 = 0.$$

Use both the fundamental theorem of algebra and the fundamental theorem of the calculus to prove (characteristic zero) that a polynomial has repeated roots if and only if the polynomial and its derivative share a root. The discriminants are found by eliminating the variable between the polynomial and its derivative.

Define the quadratic discriminant set

$$\{(a, b, c) \in \mathbf{R}^3 | b^2 - 4ac = 0\},$$

and the cubic discriminant set

$$\{(a, b, c, d) \in \mathbf{R}^4 | b^2c^2 - 4ac^3 - 4b^3d + 18abcd - 27a^2d^2 = 0\}.$$

The quadratic discriminant set is a double cone; this may not be obvious. The cubic discriminant set has even less obvious geometry. The quadratic case will become obvious below and the cubic case is worth a sculpture.

A polynomial with real coefficients could have any real root but complex roots must occur in complex conjugate pairs: zero and the real coefficients are fixed by complex conjugation. The number of real roots and even number of complex roots and the number of repeated roots is called the root type of the polynomial.

The group $GL(2, \mathbf{R})$ was defined to act on binary forms, not polynomials in one variable. For the inhomogeneous polynomials in one variable

$$ax^2 + bx + c, \qquad ax^3 + bx^2 + cx + d,$$

replace x by x/y and multiply through by the right power of y to get the homogeneous binary forms in two variables

$$ax^2 + bxy + cy^2, \qquad ax^3 + bx^2y + cxy^2 + dy^3.$$

It now makes sense to talk about repeated roots, discriminants, and root types of these binary forms.

The root type is invariant under any invertible *real* linear transformation of the two variables (x, y). If two cubic forms have the same root type there is a matrix in $GL(2, \mathbf{R})$ which sends the roots of one into the roots of the other. The corresponding substitution changes the coefficients of one into the other. Part of this picture is that the cubic can be solved in terms of cube roots and square roots involving the discriminant which is an invariant up to scale of the $GL(2, \mathbf{R})$ action. Details of this proof, that root type is invariant, are omitted. What is important here is that the root type in the quadratic and cubic case characterizes the $GL(2, \mathbf{R})$ orbits. This is not true for higher degree forms.

Root types of the quadratic and cubic forms have the special names given in the first column of the following tables.

name	root type	example quadratic	dimension
hyperbolas	two complex	$x^2 - y^2$	3
ellipses	two real distinct	$x^2 + y^2$	3
parabolas	two real equal	x^2	2
origin	—	—	0

TABLE I.

name	root type	example cubic	dimension
hyperbolic umbilics	two complex, one real	$x^3 + y^3$	4
elliptic umbilics	three real distinct	$x^3 - xy^2$	4
parabolic umbilics	three real, two equal	x^2y	3
exceptional	three real equal	x^3	2
origin	—	—	0

TABLE II.

The names come from setting the form equal to a constant and naming the corresponding curve. For example, $\{(x, y) | x^2 - y^2 = 1\}$ is a hyperbola [Golubitsky].

Transform the $(a, b, \ldots)$ according to the linear substitution

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \alpha x + \beta y \\ \gamma x + \delta y \end{pmatrix}$$

and substitute them back into the discriminant. There is a surprising simplification; the new discriminant factors into the old discriminant and a power of the determinant of the 2×2 matrix. The quadratic discriminant becomes

$$(b^2 - 4ac)(\alpha\delta - \beta\gamma)^2,$$

and the cubic discriminant becomes

$$-(b^2c^2 - 4ac^3 - 4b^3d + 18abcd - 27a^2d^2)(\alpha\delta - \beta\gamma)^6.$$

The geometric interpretation is that the discriminant sets, defined by setting the discriminant equal to zero, don't change under $GL(2, \mathbf{R})$ action. This group invariance will reveal the geometric structure.

Construct the complex numbers $\mathbf{C}$: select exactly one of the square roots of negative one, call it i, adjoin it to the real numbers $\mathbf{R}$ so that $\mathbf{C} = \mathbf{R} + i\mathbf{R}$. All complex numbers correspond to some 2×2 real matrices by

$$\alpha + \beta i \leftrightarrow \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}, \qquad 1 \leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad i \leftrightarrow \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Complex number addition, subtraction, multiplication, and division is now 2×2 matrix addition, subtraction, multiplication, and multiplication by an inverse. The imaginary i with $i^2 = -1$ is real as a 2×2 matrix whose square is the negative identity matrix. The matrix transpose corresponds to the complex conjugate. The matrix determinant is the square of the absolute value, so the multiplicative group of non-zero complex numbers, $\mathbf{C}^\times$, lies in $GL(2, \mathbf{R})$.

$\mathbf{C}^\times$ in polar rather than cartesian coordinates is

$$\alpha + \beta i \leftrightarrow \rho e^{i\theta}, \qquad e^{i\theta} \leftrightarrow \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

where we will assume that ρ is positive.

There is a complex number representation for the real binary forms. Let $z = x + yi$, the complex conjugate $z^* = x - yi$, and $u = u_1 + u_2 i$, $v = v_1 + v_2 i$. The real part of the complex quadratic form $(uz^2 + vzz^*)$ is

$$\begin{aligned} \mathbf{Re}(uz^2 + vzz^*) &= (u_1 + v_1)x^2 + -2u_2xy + (-u_1 + v_1)y^2 \\ &= ax^2 + bxy + cy^2, \end{aligned}$$

where we take v to be real, $v_2 = 0$. The real part of the complex cubic form $(uz^3 + vz^2z^*)$ is

$$\begin{aligned} &\mathbf{Re}(uz^3 + vz^2z^*) \\ &\quad = (u_1 + v_1)x^3 + (-3u_2 - v_2)x^2y + (-3u_1 + v_1)xy^2 + (u_2 - v_2)y^3 \\ &\quad = ax^3 + bx^2y + cxy^2 + dy^3. \end{aligned}$$

Comparing coefficients gives two invertible linear transformations which relate the $(a, b, \dots)$ and the (u, v) coefficient sets:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -2 & 0 \\ -1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ v_1 \end{pmatrix},$$

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & -3 & 0 & -1 \\ -3 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ v_1 \\ v_2 \end{pmatrix}.$$

These transformations simplify the discriminant sets. Look at the new form the

discriminants take. The quadratic discriminant is 4 times

$$u_1^2 + u_2^2 - v_1^2$$

and the cubic discriminant is 4 times

$$27(u_1^2 + u_2^2)^2 - (v_1^2 + v_2^2)^2 - 18(u_1^2 + u_2^2)(v_1^2 + v_2^2) + 8u_1v_1(v_1^2 - 3v_2) + 8u_2v_2(3v_1^2 - v_2^2).$$

The quadratic discriminant is obviously a double cone from the emergence of the circles $u_1^2 + u_2^2$. The cubic discriminant needs more work. First look at the quadratic discriminant another way.

Recall that the discriminant is zero when the polynomial and its derivative both vanish. The forms are homogeneous, so rescale the complex variable z by a variable of absolute value 1, $z = e^{i\varphi}$. Then

$$\begin{aligned}\mathbf{Re}(uz^2 + vzz^*) &= \mathbf{Re}(ue^{2i\varphi} + v)\\ &= v_1 + \frac{1}{2}(ue^{2i\varphi} + u^*e^{-2i\varphi}).\end{aligned}$$

Set the form

$$v_1 + \frac{1}{2}(ue^{2i\varphi} + u^*e^{-2i\varphi}) = 0$$

and the derivative

$$2iue^{2i\varphi} - 2iu^*e^{-2i\varphi} = 0.$$

Add a quarter of the second equation to the first and solve for $u = -v_1e^{2i\varphi}$. In $(u,v) = (u_1 + iu_2, v_1)$ coordinates, the quadratic discriminant is the set of $(v_1e^{i\theta}, v_1)$ for v_1 and θ real and arbitrary. These pairs obviously form a double cone: think of v_1 as a vertical axis and the $v_1e^{i\theta}$, θ arbitrary, as a circle of radius v_1 located in a plane perpendicular to the v_1 axis at a directed distance of v_1. Such a circle consists of parabolas or forms with two repeated roots. This circle is a quadratic analog of the cubic *Umbilic Torus*.

Apply this pattern of calculation to the real part of the complex cubic form

$$\begin{aligned}\mathbf{Re}(uz^3 + vz^2z^*) &= \mathbf{Re}(ue^{3i\varphi} + ve^{i\varphi})\\ &= \frac{1}{2}(ue^{3i\varphi} + u^*e^{-3i\varphi} + ve^{i\varphi} + v^*e^{-i\varphi}).\end{aligned}$$

Set the form

$$\frac{1}{2}(ue^{3i\varphi} + u^*e^{-3i\varphi} + ve^{i\varphi} + v^*e^{-i\varphi}) = 0$$

and its derivative

$$\frac{1}{2}(3iue^{3i\varphi} - 3iu^*e^{-3i\varphi} + ive^{i\varphi} - iv^*e^{-i\varphi}) = 0.$$

Add six times the first equation to the second to get

$$4ue^{3i\varphi} - 2u^*e^{-3i\varphi} + 2ve^{i\varphi} = 0.$$

Set $u = 1$, this is really two equations $u_1 = 1$ and $u_2 = 0$. Solve for v in terms of φ to get a curve in the $u = 1$ plane of (u, v) space,

$$v = -2e^{2i\varphi} + e^{-4i\varphi},$$

Change the variable $\varphi = \frac{1}{2}(\phi - \pi)$ to simplify

$$v = 2e^{i\phi} - e^{-2i\phi}, \qquad 0 \leqslant \phi < 2\pi.$$

This classical curve is the locus of a point on a circle of radius 1 rolling inside a circle of radius 3, a hypocycloid of three cusps [Brieskorn]. In the $u = 1$ plane this curve consists of parabolic umbilics, two or three repeated roots. All three roots repeated is the subset of the curve given by the second derivative of the form vanishing, or equivalently,

$$\frac{dv}{d\phi} = 2ie^{i\phi} - 2ie^{-2i\phi} = 0,$$

or

$$e^{3i\phi} = 1 = e^{2i\pi k}, \; v = 3e^{2i\pi k/3}, \qquad k \in \mathbf{Z}.$$

These scaled cube roots of unity are cusps on the curve; a tangent line but no tangent circle is defined at a cusp.

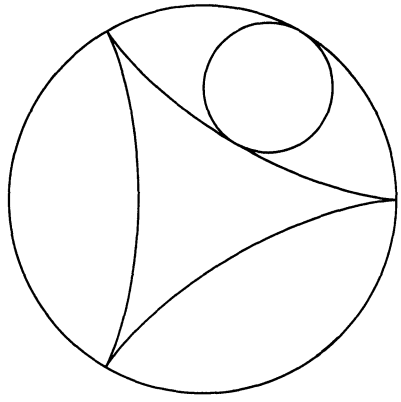


FIG. 11. Hypocycloid of Three Cusps with Generating Circles.

The group $\mathbf{C}^\times$ contains the rotation $e^{i\theta}$ which acts simply on the cubic form. Substitute $z \mapsto e^{i\theta}z$, in the real part of the complex cubic form:

$$\mathbf{Re}(uz^3 + vz^2z^*) \mapsto \mathbf{Re}(ue^{3i\theta}z^3 + ve^{i\theta}z^2z^*).$$

Geometrically

$$(u, v) \mapsto (ue^{3i\theta}, ve^{i\theta}),$$

u gets rotated thrice while v is rotated once. This group action leaves the cubic discriminant unchanged because the determinant of $e^{i\theta}$ as a matrix is exactly 1.

The plane, given by $u = 1$ or the v-plane, gets moved by the unit circle group. Recall the u thrice while v once, or one-third rotation of v while u rotates once. This means that the hypocycloid rotates one cusp over while $u = e^{0i\pi} = 1$ moves

once around and back to $u = e^{2i\pi} = 1$. The hypocycloid has three fold symmetry, so the resulting surface closes and forms a torus. There are two ways to go around a torus. The curve of cusps goes around the torus three times the long way and once the short way; the long way is defined by the rotation of the $u = 1$ plane.

Now look at this twisted torus with a hypocycloid cross-section from the point of view of the $(1, v)$ geometry. The parameterized surface has the complex exponential form

$$(\phi, \theta) \mapsto (e^{3i\theta}, 2e^{i(\phi+\theta)} - e^{i(\theta-2\phi)}), \qquad 0 \leqslant \phi, \theta < 2\pi.$$

This torus gives only one view of the cubic discriminant set. This view is the intersection of the three dimensional parabolic umbilic set or cubic discriminant set with the three-dimensional space formed by rotating the $u = 1$ plane. There are interesting intersections with other three-dimensional spaces.

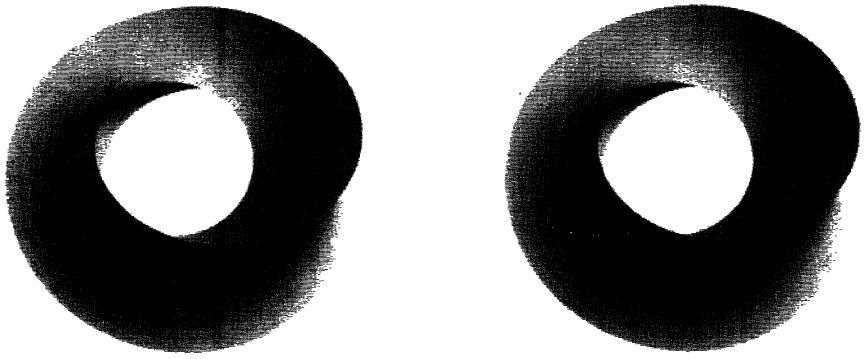


FIG. 12. Stereo Pair of $GL(2, \mathbf{R}) \times \mathbf{R}^4$: The Intersection of The Cubic Discriminant Set and a Rotated Plane.

Direct carving $GL(2, \mathbf{R}) \times \mathbf{R}^4$, new style. About the same time I was direct carving the *Torus with Cross-Cap and Vector Field* I learned how jet engine parts are made. I thought there might be some application to my sculpture. I had been thinking mathematically about sculptural forms and real objects for some years. It turned out that I had independently addressed and solved some problems that have industrial applications. The major difference between my sculpture and industrial parts is that superficially an aesthetic object tends to be more complex and interesting than a functional part. The aesthetic artifact is intended to capture human attention and become part of a social event. That aspect may be irrelevant or dangerous in machine parts. But the problems for sculpture were more difficult than a typical machine part. To solve a problem, find a harder related problem and solve that; then the first problem will be easier. So I lectured a room full of engineering faculty and graduate students for three semesters on my approach to computer-aided design and manufacture. My motivation centered around automating direct carving.

Direct carving $GL(2, \mathbf{R}) \times \mathbf{R}^4$ involved thousands of computer driven milling machine operations. Computer milling was done to preserve some of the quantitative as well as the qualitative features of the mathematics behind the fine art.

Direct carving new style by computer drive is neither better, than nor worse than direct carving old style by hand-eye-tool-medium coordination. The computer driven milling machine is a new tool, some things are easier, some things are much harder. To work consistently to the nearest thousandths of an inch does offer new options. I am satisfied that direct carving new style has expressive and aesthetic potential.

A lot of work went into revealing this twisted torus with hypocycloid cross-section. It has some significance in the theory of real binary cubic forms. We could just make a model of this torus, as exactly as physically possible. When we got done we would have just that, a model. Nothing more, something less. A model is less than what it attempts to be, a sculpture is more. A sculpture can have perhaps, nuance, timbre, mystery, warmth, history, numerous levels of meaning, and many references other than what its humble origins would hope or explain. A pure model would have no trace of its construction, some people might like that. But that is not what I did at all. I am interested in the adventure of affirming pure mathematical thought in unpredictable physical form.

Besides that I had in mind direct carving this umbilic torus. Direct carving means I move a tool around on the surface, a real physical tool leaves interesting marks, texture. The cutting tool has finite width, I have to live with this finite width, exploit the tell-tale marks of a tool that can visit only one point at a time. A tool-path is a curve. Surface filling curves can visit one point at a time and cover the entire surface. Because the cutting tool has finite width, I decided to use some order of a surface filling curve [Hilbert]. What manufacturing engineers typically do for tool-paths is have the cutter go back and forth, from one side to the other, or make a spiral motion, usually amounting to a series of concentric rectangles. These are all patterns familiar to anyone who has ever mowed a lawn. Next time mow your lawn following some order of a Hilbert curve. Surface filling curve tool paths are a new process. In some situations surface filling curves are more efficient than traditional paths.

There are many surface filling curves; such fractals abound. I chose the Peano-Hilbert curve forms because when machined in slightly rough material they have overtones of Shang and Chou bronzes, of Mayan pictographic writing. The curve itself lies centered in a groove, the groove wall cuts off the neighboring

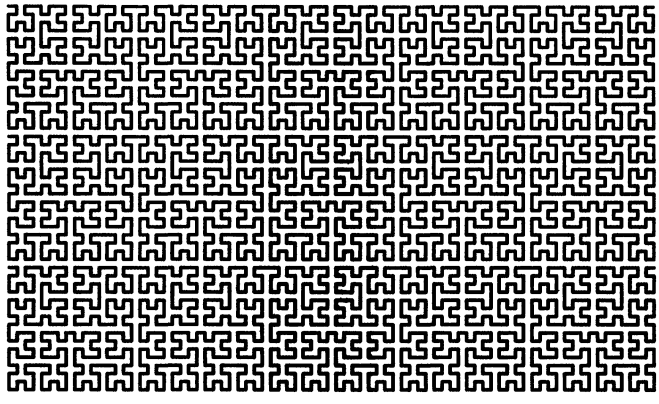


FIG. 13. Peano-Hilbert Surface Filling Curves.

groove wall to form a maze wall. These maze walls are worn down in the bronze finishing process and are interesting to look at and touch. The grooves were cut to one-third ball end mill depth, the twisted hypocycloid with three cusps already abounds in threes and thirds.

I define the nth order of the Peano-Hilbert curve recursively by constructing an ordered list of non-negative integer coordinate pairs $H[n]$. The join symbol $\sqcup$ means to preserve the order as usually read left to right and top to bottom; the addition, subtraction and matrix multiplication are supposed to thread through the list. Begin with $H[0] = (0, 0)$. Then for $n \geqslant 1$, define

$$H[n] = H[n-1] \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$\sqcup$$
$$(2^{n-1}, 0) + H[n-1]$$
$$\sqcup$$
$$(2^{n-1}, 2^{n-1}) + H[n-1]$$
$$\sqcup$$
$$(2^{n-1} - 1, 2^n - 1) - H[n-1] \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

I wanted to carve the surface of the torus of parabolic umbilics in such a way that I touched each point of the surface once. The parameters ϕ, θ lie in a square with each side of length some multiple of π. The integer coordinates of $H[n]$ grow without bound; to keep the curve in the unit square just scale by the right power of two and the right multiples of π. I say multiples because I filled my square domain with three pairs of Peano-Hilbert curves directed to go the long way around the torus. This gave the final spatial torus image more balance.

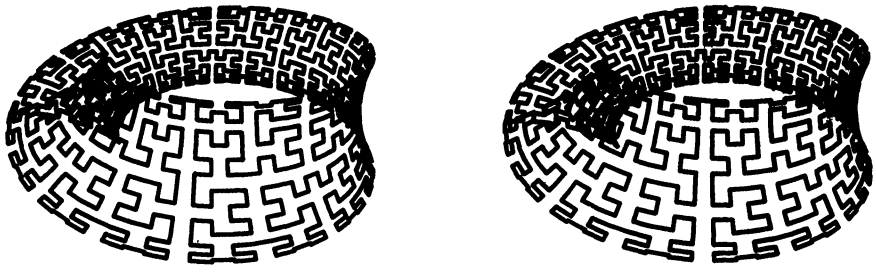


FIG. 14. Stereo Pair of the Peano-Hilbert Umbilic Torus Filling Space Curve, One Third Shown.

The cutter was a five-eighths inch ball end mill of tungsten carbide rotating at a variable rotating speed less than 10,000 rpm. I cut a high density foam positive image which was held fixed throughout the cutting process. The available computer driven milling machine was a three axis unit, the bed travel was 28 inches in the x and y directions, the cutting tool quill had a travel of twelve inches. This machine geometry meant that I must always approach my complicated surface

from a vertical direction. The tip of the tool, if allowed to follow the parabolic umbilic surface exactly would be cutting into the elliptic umbilics inside the torus. I wanted the ball end mill and the whole milling machine apparatus to consist only of hyperbolic umbilics. That is where the observers of the sculpture are, the whole exterior space comprising hyperbolic umbilics. How do you feel now that since the creation of this sculpture you are hyperbolic umbilic? To avoid cutting up elliptic umbilics a tool compensation must be made. I wanted the ball end mill to be tangent to the parabolic umbilic surface, this means that a radius vector of the sphere part of the ball end mill be normal to the parabolic umbilics. All the normals had to be calculated. This was done by writing the twisted hypocycloid torus in terms of sines and cosines, explicitly writing out two independent tangents, taking their cross-products everywhere, and normalizing those to get the unit normal vectors. A simple vector addition with the vertical drop gives the proper tool compensation.

It is sobering to debug a computer program that drives a potentially lethal tool with a thrust of over a thousand foot pound inches per second squared. For insurance many tool path pictures were drawn on a stereo imaging system. The computer languages appropriate to the computer graphics displays were inadequate for addressing the milling machine. The computer driven milling machine had its own language and file system based upon earlier paper tape. The program for the *Umbilic Torus* would have been impossible a few years earlier; the paper tape would have filled the room. As it was we had to install a hard disc drive in the milling machine.

The final cutting of the high density foam one side at a time took over eight hours. It came out looking like a most remarkable waffle. I intervened and sculpted this object to transcend the industrial process. Here was a positive image emerging in the full context of our electronic information processing era which I was about to immerse into bronze age technology. The bronze casting involved the traditional inversions. The original positive image was inverted to a negative silicon rubber mold, which was inverted to a positive hard wax image. This positive wax was dismantled and the pieces sprued with wax attachments to facilitate good liquid bronze flow; the positive results were dipped in a ceramic shell material and dried repeatedly to form a ceramic shell negative flask. This ceramic shell negative was fired in a kiln. The final inversion was when the positive bronze was poured. Fine art bronze contains some silicon and zinc with the copper (brass has tin instead of zinc). The bronze was welded together, chased, and given a corrosion arresting patina of antique greens. These bronze foundry processes date back several millenia. Some of the mathematics is ancient too; understanding both the torus and the hypocycloid calls up the ancient geometry of wheels. It was striking to see the old and the new mathematics and computer abstractions interact with the old and the new foundry and milling realizations.

The JPBM poster. The Joint Policy Board for Mathematics (JPBM), located in Washington, D.C., is a consortium of the American Mathematical Society, the Mathematical Association of America, and the Society for Industrial and Applied Mathematics. In connection with *Mathematics Awareness Week* in the Spring of 1989, the JPBM published an elegant five color poster featuring my silicon bronze sculpture *Umbilic Torus NC* [Hoffman].

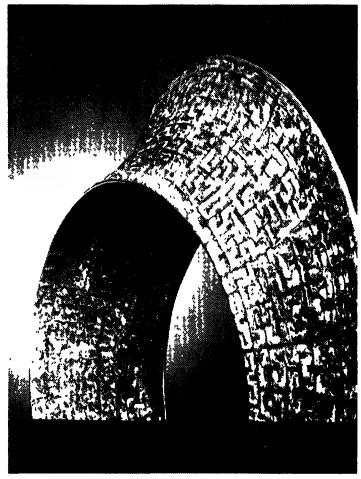


FIG. 15. *Umbilic Torus NC* Close Up, Quarter View.

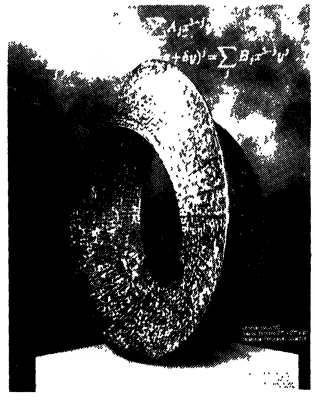


FIG. 16. *JPBM Poster Featuring Umbilic Torus NC.*

The mathematics behind *Umbilic Torus NC* includes a fundamental representation of the matrix group $GL(2, \mathbf{R})$. The 'NC' is a cutting technology abbreviation for 'numerical control'. The fundamental group representation appears in robin egg blue script in the background on the JPBM poster. We have discussed this mapping of 2×2 matrices into 4×4 matrices via linear substitutions on binary cubic forms, e.g.,

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}: \sum_{0 \leqslant j \leqslant 3} A_j x^{3-j} y^j \mapsto \sum_{0 \leqslant j \leqslant 3} A_j (\alpha x + \beta y)^{3-j} (\gamma x + \delta y)^j = \sum_{0 \leqslant j \leqslant 3} B_j x^{3-j} y^j.$$

The only difference between this expression and that on the poster is the obscure point that the range of summation there is simply given over j whereas it is to be understood that $0 \leqslant j \leqslant 3$.

Acknowledgements. Relative to *Torus with Cross-Cap and Vector Field* I thank Brent Gehring of the Brigham Young University Sculpture Department for the space, compressed air, hoist, and reasonable price on the marble, Peter Crawley of the BYU Mathematics Department for reminding me of his catalytic cross-cap

paper at just the right time, Don Robinson of the same department and the MAA Council for facilitating the MAA gift to the AMS, and Richard Askey of the University of Wisconsin for suggesting that sculpture be part of the AMS Centennial celebration. Relative to *Umbilic Torus NC* I thank Jordan Cox of the Brigham Young University Mechanical Engineering and Technology Departments for introducing me to industrial grade cutting technology, Kent Kohkonen for the run of the BYU Robotics and Automated Manufacturing Laboratory, BYU Mechanical Engineering Graduate Student Paul Oliphant for his brilliant computer programming, Professor Chris Zeeman of Warwick University and Tim Poston for discussions at Battelle Seattle on the cubic discriminant, Steve Wolfram for organizing Mathematica(TM), and Neil Hadlock of Wasatch Bronzeworks for providing the lost wax ceramic shell bronze casting facility.

BIBLIOGRAPHY

Vladimir Igorevich Arnold, *Catastrophe Theory*, second, revised and expanded Edition, translated from the Russian by Gordon S. Wassermann and R. K. Thomas, page 90 for the quote, Springer-Verlag, New York, 1986, xiii + 108 pages.

Egbert Brieskorn and Horst Knörrer, *Plane Algebraic Curves*, translated from the *Ebene algebraische Kurven* by John Stillwell, Birkhäuser Verlag, Basel, 1986, vi + 721 pages.

C. E. Burgess, *Classification of Surfaces*, American Mathematical Monthly, Volume 92, Number 5, May 1985, pages 349–354.

James W. Cannon, *Mathematics in Marble and Bronze: the Sculpture of Helaman Rolfe Pratt Ferguson*, Mathematical Intelligencer, to appear 1990, 16 reproductions.

Arthur Cayley, *On the Geometrical Interpretation of the Covariants of a Binary Cubic*, Collected Mathematical Papers, volume 7, pages 332–333, Cambridge University Press, 1894, reprinted from the Quarterly Journal of Pure and Applied Mathematics, volume x, 1870, pages 148–149.

Peter Crawley, *On the Equivalence of Two Surfaces*, American Mathematical Monthly, Volume 67, Number 2, February 1960, pages 165–166.

Helaman R. P. Ferguson, *Algorithms, Geometry, and GL(n, Z)*, MAA Invited Hour Address, American Mathematical Society and Mathematical Association of America Joint Meetings, Anaheim, California, January 1985.

Helaman R. P. Ferguson, *A Noninductive GL(n, Z) Algorithm That Constructs Integral Linear Relations for n Z-Linearly Dependent Real Numbers*, Journal of Algorithms, Volume 8, 1987, pages 131–145.

Helaman Ferguson, *'Torus with Cross-Cap and Vector-Field'*, White Carrara Marble, 545 lbs, $3' \times 2\frac{1}{2}' \times 1\frac{1}{2}'$, this sculpture was a gift from the Mathematical Association of America to the American Mathematical Society on the occasion of its Centennial, was permanently installed January 1988, and can be seen in the American Mathematical Society National Headquarters Building, 201 Charles Street, Providence, Rhode Island 02940.

Helaman Ferguson, *Umbilic Torus NC*, Silicon Bronze, dimensions $27'' \times 27'' \times 9''$, base $15'' \times 15'' \times 34''$, a limited edition of twelve lost wax ceramic shell bronze castings. For example, number four of this edition was a gift from Professor William O. Reinhardt, M.D., formerly dean of the Medical School, to the Medical Center Library, University of California at San Francisco, Parnassus Way, San Francisco, California 94143.

Martin Golubitsky and Victor Guillemin, *Stable Mappings and Their Singularities*, Chapter VII, Classification of Singularities, Part II: The Local Ring of a Singularity, §Umbilic, pages 182–190, Graduate Texts in Mathematics, Volume 14, Springer-Verlag, New York, 1973, x + 209 pages.

David Hilbert, *Über die stetige Abbildung einer Linie auf ein Flächenstück*, Mathematische Annalen, Band 38, 1891, Seiten 459–460.

Kenneth Hoffman, Kirsten Sampson, Kathrine Molloy, *Mathematics: Discovering Patterns*, Five Color Poster $16'' \times 20''$, featuring *Umbilic Torus NC*, Silicon Bronze, dimensions $27'' \times 27'' \times 9''$, also the two full color postcards: *Alexander's Horned Wild Sphere*, *Five Bifurcations*, Silicon Bronze, dimensions $12'' \times 12'' \times 5''$; *Torus with Cross Cap*, Silicon Bronze, dimensions $10'' \times 3'' \times 7''$, sculpture by Helaman Ferguson, photography by Terry Clough, poster design by Karen O'Brien Winkler; published by the Joint Policy Board for Mathematics, Office of Governmental and Public Affairs, 1529 18th Street N.W., Washington, D.C. 20036, 202-234-9570, May 1989, edition of 5,000 each, less than 600 posters left, less than 50 postcards each left.

Allyn Jackson, *The AMS Centennial 1888–1988*, Notices of the American Mathematical Society, January 1988, Volume 35, Number 1, pages 15–20.

David Mumford, *Geometric Invariant Theory*, The Hoary Problem of Binary Quantics, page 77, Springer-Verlag, 1965.

Everett Pitcher, James Maxwell, Hope Daly, Patricia Cross, Edith Krekorian, *AMS Centennial 1888–1988*, Four Color Poster, featuring *Torus with Cross-Cap and Vector-Field*, White Carrara Marble, 545 lbs, $3' \times 2\frac{1}{2}' \times 1\frac{1}{2}'$, sculpture by Helaman Ferguson, photography by Mark Philbrick, poster design by Susan Northrop, published by the American Mathematical Society, 201 Charles Street, Providence, R.I. 02940, 401-455-4000, January 1988, edition of 5,000, less than 900 left.

Itsuo Sakane, *Heso no o-Umbilic Torus*, Kaleidoscope of Form, The Natural Poetry of Science and Art, Kagaku Asahi-Monthly Science Magazine, Asahi Shimbun, October, 1989, six full color reproductions, pages 5, 70–72.

Peter Scott, *The Geometries of 3-Manifolds*, Bulletin of the London Mathematical Society, volume 15, 1983, pages 401–487.

William P. Thurston, *The Geometry and Topology of 3-Manifolds*, Princeton University Notes, 1978–1979, Nine Chapters.

Eric Christopher Zeeman, *The Umbilic Bracelet and the Double-Cusp Catastrophe*, pages 331–338, in Peter Hilton, editor, Structural Stability, the Theory of Catastrophes, and Applications in the Sciences, Battelle Seattle Research Center, 1975, Lecture Notes in Mathematics, Volume 525, Springer-Verlag, New York, 1976, 408 pages.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, FREDERICK W. LUTTMAN, MARVIN MARCUS, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.

*An asterisk (*) after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before February 28, 1991 at the MONTHLY PROBLEMS address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.

ELEMENTARY PROBLEMS

E 3397. *Proposed by Ji Chen and Cheng-hui Lo, University of Science and Technology, Hefei, Anhui, People's Republic of China.*

The perimeter of a triangle ABC is divided into three equal parts by three points P, Q, R. Show that

$$\text{Area}(\triangle PQR) > \frac{2}{9}\text{Area}(\triangle ABC)$$

and that the constant $2/9$ is best possible.

E 3398. *Proposed by Alan H. Stein, University of Connecticut, Waterbury.*

Find all pairs of positive integers m, n such that $\phi(m)|n$ and $\phi(n)|m$, where ϕ denotes Euler's function.

E 3399. *Proposed by Robert W. Floyd, Stanford University, CA.*

On a certain island with n married couples, every couple consists of a hunter and a farmer. The Ministry of Hunting has divided the island into n hunting ranges of equal size A. Working independently, the Ministry of Agriculture has divided the island into n farming ranges of equal size A. The Ministry of Marriage insists that the hunting range and the farming range assigned to each couple be close together. To everyone's surprise the Ministry of Assignments is able to allot the hunting ranges and the farming ranges to the various couples in such a way

that each couple's two ranges overlap. The Ministry of Religion declares this to be a miracle.

(a) Show that in fact no miracle is involved by proving the existence of a positive number δ_n depending only on n with the following property: No matter how the Ministries of Hunting and Agriculture have made their divisions, it is possible for the Ministry of Assignments to make its choices in such a way that each couple's two ranges overlap by an area at least $\delta_n A$.
(b) Determine the best possible value of δ_n.

E 3400. *Proposed by Burt J. Totaro, Mathematical Sciences Research Institute, Berkeley, CA.*

Let S be the boundary of the unit square $[0, 1] \times [0, 1]$ in R^2. Suppose f is a continuous real-valued function on S such that $f(x, 0)$ and $f(x, 1)$ are polynomial functions of x on $[0, 1]$ and such that $f(0, y)$ and $f(1, y)$ are polynomial functions of y on $[0, 1]$. Prove that f is the restriction to S of a polynomial function of x and y.

E 3401. *Proposed by James A. Davis, Michael Kerckhove, and J. Van Bowen, University of Richmond, VA.*

Suppose n points are independently chosen at random on the perimeter of a circle. What is the probability that all the points lie in some semicircle?

E 3402. *Proposed by Joseph Kupka, Monash University, Clayton, Victoria, Australia.*

A population consisting of particles of various types evolves in time according to the following rule: Each particle is deemed to belong to a unique generation $n = 1, 2, 3, \ldots$. Each particle produces a certain number of "offspring" particles, and, for each n, generation $n + 1$ comprises the totality of offspring of the particles in generation n. A particle of type $i = 0, 1, 2, \ldots$ produces exactly $i + 2$ offspring, one each of types $0, 1, 2, \ldots, i + 1$. Let $N(n, k)$ denote the number of particles in the nth generation when the first generation consists of a single particle of type k. Find a formula for $N(n, k)$.

SOLUTIONS OF ELEMENTARY PROBLEMS

A New Property of an Old Configuration

E 3251 [1988, 131]. *Proposed by Clark Kimberling, University of Evansville, Evansville, IN.*

Given a triangle ABC, it is possible to construct three circles $\Gamma_A, \Gamma_B, \Gamma_C$ inside ABC in such a way that each one is externally tangent to the other two, Γ_A is tangent to the two sides AB and AC, Γ_B is tangent to the two sides BC and BA, and Γ_C is tangent to the two sides CA and CB. (Cf. A. S. Winsor, *Modern Higher Plane Geometry*, Christopher Publishing House, Boston, 1941, pp. 206–209 or H. G. Forder, *Higher Course in Geometry*, pp. 244–245.) Let D be the point of

contact of Γ_B and Γ_C, E be the point of contact of Γ_C and Γ_A, and F be the point of contact of Γ_A and Γ_B.

(a) Prove that the lines AD, BE, and CF are concurrent.

*(b) Prove that the lines $A'D$, $B'E$, and $C'F$ are concurrent, where A', B', and C' are the excenters of the triangle ABC.

Solution by I. G. Macdonald, Queen Mary College, London, England. (a) Given three circles in a plane, each pair has two centers of similitude, and four triples among the resulting six points are collinear. In the case of the circles $\Gamma_A, \Gamma_B, \Gamma_C$ with centers X, Y, Z, three of these six points are D, E, F, and the remaining three, D', E', F', are the points of intersection of the external common tangents of Γ_B and Γ_C, Γ_C and Γ_A, and Γ_A and Γ_B, respectively. The points $D'EF$ lie on a line p, the points $DE'F$ on a line q, and the points DEF' on a line r, and finally $D'E'F'$ on a line s. For $D'EF$, collinearity depends on the fact that the line EF makes equal angles with the circles Γ_B and Γ_C and, therefore, passes through their center of similitude D'; similarly for $DE'F$ and DEF'. Collinearity of the points $D'E'F'$ follows from a theorem of Monge: the three intersection points of common external tangents for three circles taken two at a time are collinear (see Hilbert and Cohn-Vossen, *Geometry and the Imagination*, Chelsea, 1952, p. 137).

Meanwhile, the centers of similitude are collinear with the centers of a pair of circles, so we have collinear sets $DD'YZ$, $EE'ZX$, and $FF'XY$. We have D' on the three lines EF, YZ, BC, E' on the three lines FD, ZX, CA, and F' on the three lines DE, XY, AB. Hence the three triangles DEF, XYZ, ABC have s as a common axis of perspective. Since the pair of triangles DEF, ABC is in perspective, the lines AD, BE, CF meet in a common point, as desired.

(b) We shall use the theorem that two triangles in a plane are in perspective if and only if there is a conic with respect to which they are reciprocal, meaning that the sides of each triangle are the polars of the vertices of the other (see, for example, H. F. Baker, *Introduction to Plane Geometry*, Cambridge Univ. Press (1943), p. 191). Since the triangles ABC, XYZ are in perspective, there is a conic Σ with respect to which they are reciprocal. Let P, Q, R, S be the poles with respect to Σ of the lines p, q, r, s, respectively.

We next show that S is the incenter and P, Q, R are the excenters A', B', C' of the triangle ABC. Indeed, since s is the axis of perspective of ABC and XYZ, the point S is their center of perspective. But AX, BY, CZ are the internal angle-bisectors of the triangle ABC, so S is also the incenter of ABC. Moreover, since XYZ is the diagonal triangle of the quadrilateral $pqrs$, it follows that ABC is the diagonal triangle of the quadrangle $PQRS$. Hence the lines PS and QR pass through A and are harmonically separated by AB and AC. Since AS is the internal bisector of the angle A, it follows that QR is the external bisector. Likewise RP and PQ are the external bisectors of the angles B and C, and therefore P, Q, R are the excenters of the triangle ABC.

By construction, the triangles DEF (with sides on p, q, r) and PQR are reciprocal with respect to Σ. Hence they are in perspective, by the theorem quoted above. This means that the lines DP, EQ, FR are concurrent, as desired.

Solved also by E. Raytchev. Part (a) was solved by J. Dou (Spain) and the proposer. Two incorrect solutions were received.

A Binomial Coefficient Inequality

E 3309 [1989, 154]. *Proposed by Donald E. Knuth, Stanford University, Stanford, CA.*

Suppose that a, b, and c are nonnegative integers with $b > a$. If $0 < \theta < 1$ and $0 \leqslant x \leqslant 1$, prove that

$$(-1)^{b-a-1} \sum_{k=0}^{c} \binom{c}{k}\binom{a+\theta}{b+k} x^k > 0.$$

Here (as usual)

$$\binom{t}{k} = t(t-1)\cdots(t-k+1)/k!$$

for any real t and positive integer k, while $\binom{t}{0} = 1$.

Solution I by David Callan, University of Bridgeport, Bridgeport, CT. Fix a and θ and let $F(x; b, c)$ denote the left-hand side of the given inequality. We show by induction on c that for any integer $b > a$, $F(x; b, c) > 0$. For $c = 0$ or $x = 0$, $F(x; b, c) = (-1)^{b-a-1}\binom{a+\theta}{b}$, which is easily seen to be positive. Replacing k by $c - k$ in the sum, we can use the Vandermonde convolution $\Sigma_k \binom{r}{k}\binom{s}{t-k} = \binom{r+s}{t}$ to obtain $F(1; b, c) = (-1)^{b-a-1}\binom{a+c+\theta}{b+c} > 0$. One can directly verify that $F'(x; b, c) = -cF(x; b+1, c-1)$. Together these facts provide a basis for the induction.

Solution II by the editors. The sum is easily seen to be equal to the coefficient of u^{b+c} in $(-1)^{b-a-1}(u+x)^c(u+1)^{a+\theta}$. Expanding $(u+x)^c = [(u+1)+(x-1)]^c$ in powers of $u+1$, we see that the sum is equal to

$$\sum_{i=0}^{c} (-1)^{b-a-1}\binom{c}{i}\binom{c+a+\theta-i}{b+c}(x-1)^i.$$

It is easily verified that each nonzero term is positive. (When $x = 1$ there is only one nonzero term.)

Editorial comment. Many solvers noted that the sum could be expressed in terms of the hypergeometric series as

$$(-1)^{-a-1}\frac{(-a-\theta)_b}{b!}\,{}_2F_1\left(\begin{matrix}-c, b-a-\theta\\ b+1\end{matrix}; x\right),$$

where $(r)_k = r(r+1)\cdots(r+k-1)$. Some solvers used known transformations for the hypergeometric series to express this as a sum of positive terms (Solution II above derives one of these transformations). Others used the integral representation for the hypergeometric series to express the sum as an integral of a positive function.

Solved also by R. Askey, D. W. Bailey, A. A. Jagers (The Netherlands), H. Ki (Korea), A. G. Kostenbauder, E. Levine, C. Libis, O. P. Lossers (The Netherlands), D. Secrest, J. H. Steelman, V. Strehl (West Germany), R. J. Wagner, D. Zeilberger, and the proposer.

Determining a Secret Number by Asking Suitable Primality Questions

E 3310 [1988, 155]. *Proposed by Albert Wilansky, Lehigh University, Bethlehem, PA.*

I have a secret positive integer u, not exceeding some predetermined bound N. If you give me any positive integer n, I will tell you whether or not $u + n$ is prime.

(a) Give a number $G(N)$ and a procedure such that you can always determine the value of u in at most $G(N)$ such trials.

(b) Give a number $F(N)$ such that no strategy will always determine u in fewer than $F(N)$ such trials.

Solution I of (a) by A. Zulauf, University of Waikato, Hamilton, New Zealand.

$G(N) \leqslant N$ will suffice. To see this we first note that there exists a prime p greater than N such that each of the numbers $p - j$ for $j = 1, 2, \ldots, N - 1$ is composite. (Cf. Niven and Zuckerman, *An Introduction to the Theory of Numbers*, Theorem 1.18. For example, if $N = 14$, we may take $p = 127$.) Suppose one now takes $n_i = p - N + i - 1$ for $i = 1, 2, \ldots, N$, so that $u + n_i = p - (N - u) + i - 1$. Then $p - N < u + n_1 \leqslant p$ and $u + n_N = p + u - 1 \geqslant p$. Let I be the first value of i for which $u + n_i$ is prime. In view of our choice of p, we must have $u + n_I = p$, so that u is given by the formula

$$u = N + 1 - I$$

Solution II of (a) by Imre Z. Ruzsa, Mathematical Institute of the Hungarian Academy of Science, Budapest. We show that $G(N) \leqslant \{1 + o(1)\}(\log N)^2$ and that for large N only test integers n_i not exceeding N need be used. We begin with the following lemma.

LEMMA. *If $1 \leqslant u < v \leqslant N$, let $A_{u,v}(N)$ be the set of positive integers n not exceeding N such that exactly one of the numbers $u + n, v + n$ is prime. Given $\varepsilon > 0$, we have*

$$|A_{u,v}(N)| > \frac{2}{1 + \varepsilon} \frac{N}{\log N}$$

for sufficiently large N uniformly in u, v.

Proof of Lemma. Clearly

$$\begin{aligned} |A_{u,v}(N)| = {} & |\{n: n \leqslant N, v + n \text{ is prime}\}| + |\{n: n \leqslant N, u + n \text{ is prime}\}| \\ & - 2|\{n: n \leqslant N, \text{both } u + n \text{ and } v + n \text{ are prime}\}|. \end{aligned} \tag{$*$}$$

By the prime number theorem the first term in $(*)$ is equal to

$$\int_v^{N+v} \frac{dt}{\log t} + O\left(\frac{N}{(\log N)^2}\right) = \frac{N}{\log N} + O\left(\frac{N}{(\log N)^2}\right)$$

uniformly in v. Similarly the second term on the right of $(*)$ is equal to $N/\log N + O(N/(\log N)^2)$ uniformly in u. Now, using Brun's upper bound sieve method, Schnirelmann proved that the third term on the right of $(*)$ is less than an

absolute constant times

$$\frac{N}{(\log N)^2}\prod_{p\mid v-u}\left(1+\frac{1}{p}\right)=\frac{N}{(\log N)^2}\sum_{d\mid v-u}\frac{\mu^2(d)}{d}\leqslant\frac{N}{(\log N)^2}\sum_{d\mid v-u}\frac{1}{d}.$$

(Cf. Satz 88 of E. Landau, *Über einige neuere Fortschritte der additiven Zahlentheorie*, Cambridge Tract No. 35, London, 1937 or Corollary 2.4.1 of H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, 1974). Further

$$\sum_{d\mid v-u}\frac{1}{d}=\frac{\sigma(v-u)}{v-u}=O(\log\log 3(v-u))=O(\log\log N).$$

(Cf. §22.9 of G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, 1960.) Hence the third term on the right of $(*)$ is $O(N(\log N)^{-2}\log\log N)$. Combining these results, we obtain

$$|A_{u,v}(N)|=\frac{2N}{\log N}+O\left(\frac{N\log\log N}{(\log N)^2}\right)$$

and the lemma is proved.

The result claimed at the beginning of our solution is embodied in the following theorem.

THEOREM. *If $\varepsilon>0$, N is sufficiently large, and $k=\lceil(1+\varepsilon)(\log N)^2\rceil$, then there exists a set $\{n_1,n_2,\ldots,n_k\}$ with $1\leqslant n_i\leqslant N$ such that a knowledge of which numbers $u+n_i$ $(i=1,2,\ldots,k)$ are prime uniquely determines u.*

Proof of Theorem. We consider all k-tuples $(n_1,n_2,\ldots,n_k)$ of distinct integers from $[1,N]$. There are $N(N-1)\cdots(N-k+1)$ such k-tuples; we shall show that a positive proportion of them give a set which satisfies the conclusion of the theorem.

If u,v are given integers satisfying $1\leqslant u<v\leqslant N$, then the number of k-tuples of integers from $[1,N]\setminus A_{u,v}(N)$ is at most

$$\left(N-\frac{2}{1+\varepsilon}\frac{N}{\log N}\right)^k$$

The number of pairs u,v satisfying $1\leqslant u<v\leqslant N$ is $N(N-1)/2$. Thus the total number of k-tuples from $[1,N]$ each of which avoids at least one of the sets $A_{u,v}(N)$ does not exceed

$$\begin{aligned}\frac{N(N-1)}{2}\left(N-\frac{2}{1+\varepsilon}\frac{N}{\log N}\right)^k&<\frac{1}{2}N^{k+2}\left\{1-\frac{2}{(1+\varepsilon)\log N}\right\}^k\\&<\frac{1}{2}N^{k+2}\exp\left\{\frac{-2k}{(1+\varepsilon)\log N}\right\}\\&\leqslant\frac{1}{2}N^k\\&<\frac{2}{3}N(N-1)\cdots(N-k+1),\end{aligned}$$

since

$$\frac{N(N-1)\cdots(N-k+1)}{N^k} > \frac{(N-k)^k}{N^k} = \left(1-\frac{k}{N}\right)^k > 1-\frac{k^2}{N} > \frac{3}{4}$$

for sufficiently large N. Thus there exists a k-tuple of distinct integers from $[1, N]$ (in fact there are many such k-tuples) which does not avoid *any* of the sets $A_{u,v}(N)$, i.e., which contains a representative of each $A_{u,v}(N)$ for $1 \leqslant u < v \leqslant N$. And if n_i is in $A_{u,v}(N)$, then exactly one of the integers $u + n_i, v + n_i$ is prime, so that the ith trial distinguishes between u and v. Thus the theorem is proved.

Solution of (b) by Albert Nijenhuis, Seattle, WA. We claim that $F(N) \geqslant [\log N/\log 2]$. For there are 2^k possible outcomes to a set of k trials and if these k trials determine u in each of the N cases, we must have $2^k \geqslant N$ or $k \geqslant \log N/\log 2$.

Editorial comment. Ruzsa's solution of (a) improves upon the shorter solution of Zulauf in two respects: The estimate of $G(N)$ is much sharper and the values of the test integers n are much smaller.

If we use only a qualitative form of the lemma in Ruzsa's solution, we see that, for large N, the value of u is uniquely determined by a knowledge of which of the numbers $\{u + 1, u + 2, \ldots, u + N\}$ are primes. We know of no proof of this fact which does not use sieve methods in some form. It is not obvious how large N must be for this assertion to be true; certainly we must have $N > 4$, since $n = 5$ is the smallest positive integer for which precisely one of the integers $2 + n, 4 + n$ is prime.

On the other hand Ruzsa's solution is an existence proof and does not give an explicit procedure for finding a set of test integers n_i, although an examination of only $\exp(\{1 + 0(1)\}\{\log N\}^3)$ different subsets of $\{1, 2, \ldots, N\}$ is needed. Even for modest values of N, it is hard to find the smallest possible set $\{n_1, n_2, \ldots, n_k\}$ such that a knowledge of which integers in $\{u + n_1, u + n_2, \ldots, u + n_k\}$ are prime determines u uniquely. When $N = 50$, say, such a minimal set must contain an element from each of the 1225 sets $A_{u,v}(50)$ with $1 \leqslant u < v \leqslant 50$; this selection process is rather delicate, since some of these sets $A_{u,v}(50)$ are rather small, e.g., $A_{20,50}(50) = \{27, 29, 41\}$. For $N = 5$ a minimal set of test integers is given by $n = 1$, $n_2 = 4$, $n_3 = 5$, since $A_{1,3}(5) = \{1\}$, $A_{1,4}(5) = \{5\}$, and $A_{3,5}(5) = \{4\}$.

Solved by A. Nijenhuis, I. Z. Ruzsa, A. Zulauf, and the editors.

Conditions for a Trivial Monoid

E 3311 [1989, 155]. *Proposed by Sydney Bulman-Fleming and Kenneth McDowell, Wilfrid Laurier University, Waterloo, Ontario.*

Suppose S is a monoid containing elements a and b such that every element of $S \times S$ is of the form (au, bu) for some u in S (i.e., $S \times S$, considered as a right S-set, is cyclic).

(a) Show that S must be a singleton if S is any one of the following: finite, commutative, idempotent, or inverse.

(b) Show that S need not be a singleton in general.

Solution by Jesús Ferrer, Universidad Complutense, Madrid, Spain. (a) Assume that S is a monoid that satisfies the hypothesis. We first observe that if S is either commutative, finite, idempotent, or inverse, then a admits a left inverse in S.

This observation is trivial if S is inverse. In any case, there exists $w \in S$ such that $(1, 1) = (aw, bw)$, where 1 is the identity of S. If S is commutative, then the element w is a left inverse for a. If S is finite or idempotent, then there are nonnegative integers m, n with $m > n$ and $a^m = a^n$. Now $1 = aw = a^2w^2 = \cdots$, so $1 = a^n w^n = a^m w^n = a^{m-n}a^n w^n = a^{m-n-1}a$, and a^{m-n-1} is a left inverse for a.

Let a' be a left inverse for a. If $(x, y) \in S \times S$, choose $u, v \in S$ such that $(x, y) = (au, bu)$ and $(x, x) = (av, bv)$. Then $u = a'au = a'x = a'av = v$ and $x = bv = bu = y$, and S is a singleton.

(b) Let M be the monoid of transformations from the positive integers to themselves under ordinary function composition, where $(fg)(x) = g(f(x))$. Define $a, b \in M$ by $a(k) = 2k - 1$ and $b(k) = 2k$. If $(x, y) \in M \times M$, defining $u \in M$ by $u(2k - 1) = x(k)$ and $u(2k) = y(k)$ yields $(x, y) = (au, bu)$, as desired.

Solved also by O. C. Garcia, F. W. Humburg, A. A. Jagers (The Netherlands), J.-M. Monier (France), A. Montgomery, P. G. Walsh (Canada), V. Ziegler (student), and the proposers. One anonymous solution was received.

When Fractions Contribute Nothing

E 3317 [1989, 254]. *Proposed by Peter J. Ferraro, Roselle Park, NJ.*

For any positive integer n and any positive real number α, put

$$S_n(\alpha) = \sum_{j=0}^{n-1} \left\lfloor \sqrt{\alpha + j/n} \right\rfloor.$$

(a) Prove that $S_n(\alpha) = n\lfloor\sqrt{\alpha}\rfloor$ if $\lfloor\alpha\rfloor + 1$ is not a perfect square.

(b) What is the value of $S_n(\alpha)$ if $\lfloor\alpha\rfloor + 1$ is a perfect square?

Solution by Marcin E. Kuczma, University of Warsaw, Poland. (a) Note that $S_n(\alpha) \geqslant n\lfloor\sqrt{\alpha}\rfloor$, since each summand is at least $\lfloor\sqrt{\alpha}\rfloor$. If equality does not hold, then the last summand exceeds $\lfloor\sqrt{\alpha}\rfloor$, which means that $\sqrt{\alpha} < k \leqslant \sqrt{\alpha + (n-1)/n}$ for some integer k. This requires $k^2 + 1/n \leqslant \alpha + 1 < k^2 + 1$, whence $\lfloor\alpha\rfloor + 1 = k^2$.

(b) When $\lfloor\alpha\rfloor + 1 = k^2$ for a positive integer k, we claim that the value of $S_n(\alpha)$ is $n\lfloor\sqrt{\alpha}\rfloor + \lfloor n(\alpha - \lfloor\alpha\rfloor)\rfloor$. Let $r = \lfloor n(\alpha - \lfloor\alpha\rfloor)\rfloor$; this is just a complicated way of writing

$$k^2 + r/n \leqslant \alpha + 1 < k^2 + (r+1)/n, \qquad r \in \{0, 1, \ldots, n-1\}.$$

This implies $\sqrt{\alpha + (n-r-1)/n} < k \leqslant \sqrt{\alpha + (n-r)/n}$, which is equivalent to

$$\left\lfloor \sqrt{\alpha + j/n} \right\rfloor = \begin{cases} k-1 & \text{for } j < n - r \\ k & \text{for } j \geqslant n - r. \end{cases}$$

Thus $S_n(\alpha) = (n-r)(k-1) + rk = n(k-1) + r$, which equals the formula claimed because $k - 1 = \lfloor\sqrt{\alpha}\rfloor$.

Editorial comment. W. Janous and C. Wildhagen noted that the result generalizes immediately from square roots to mth roots, and R. J. Hendel noted that the generalization goes all the way to any strictly increasing function whose inverse maps integers to integers.

Solved also by K. F. Andersen, J. C. Binz (Switzerland), W. E. Briggs, D. Callan, B. de Weger (The Netherlands), P. H. Duc (Denmark), J. Ferrer (Spain), S. Gaignoux (France), J. Ganapathy, J. Guo (China), R. J. Hendel, S. Houchen, A. A. Jagers (The Netherlands), W. Janous (Austria), P. Khajeh-Khalili, H. Ki (South Korea), N. F. Lindquist, J. B. Muskat (Israel), J. H. Nieto (Venezuela), T. S. Norfolk, J. H. Steelman, C. Wildhagen (The Netherlands), A. Zulauf (New Zealand), M. Zweiback, and the proposer.

A Best-Possible Quadratic Inequality

E 3320 [1989, 356]. *Proposed by M. S. Klamkin, University of Alberta, Edmonton, Alberta.*

Determine positive constants a and b such that the inequality

$$yz + zx + xy \geqslant a(y^2z^2 + z^2x^2 + x^2y^2) + bxyz$$

holds for all nonnegative x, y, z with $x + y + z = 1$ and is the best possible inequality of this form (in the sense that the inequality need not hold if a or b is increased).

Solution by Mark Ashbaugh, University of Missouri, Columbia, MO. The family $(a, b) = (a, 9 - a)$ for $0 < a \leqslant 4$ gives all solution pairs to this problem. Using the values $(x, y, z) = (1/3, 1/3, 1/3)$ and $(x, y, z) = (1/2, 1/2, 0)$, we obtain the two inequalities $a + b \leqslant 9$ and $a \leqslant 4$. Thus we need only prove that the desired inequality is valid for the pairs in the family described, i.e., that the inequality

$$yz + zx + xy \geqslant a(y^2z^2 + z^2x^2 + x^2y^2) + (9 - a)xyz \quad \text{for } 0 < a \leqslant 4$$

holds for nonnegative x, y, z with $x + y + z = 1$. If we let $P = (x + y + z)^2(yz + zx + xy)$, $Q = y^2z^2 + z^2x^2 + x^2y^2$, and $R = (x + y + z)xyz$, then this is equivalent to proving

$$P(x, y, z) \geqslant aQ(x, y, z) + (9 - a)R(x, y, z) \quad \text{for } 0 < a \leqslant 4,$$

for all nonnegative x, y, z, since these terms are homogeneous.

The proof is now completed by showing that $P \geqslant 4Q + 5R$ and that $Q \geqslant R$, because together these imply $P \geqslant aQ + (4 - a)Q + 5R \geqslant aQ + (9 - a)R$ for $0 < a \leqslant 4$. To obtain the first inequality $P \geqslant 4Q + 5R$ we note that

$$P = 2Q + 5R + x^3y + xy^3 + y^3z + yz^3 + x^3z + xz^3$$

and then

$$P - 4Q - 5R = xy(x - y)^2 + yz(y - z)^2 + xz(x - z)^2 \geqslant 0. \qquad (*)$$

To obtain the second inequality, $Q \geqslant R$, we apply the Cauchy-Schwarz inequality to the vectors $V = (xy, yz, xz)$ and $W = (xz, xy, yz)$ and conclude that

$$Q = \|V\| \cdot \|W\| \geqslant V \cdot W = x^2yz + xy^2z + xyz^2 = R. \qquad (**)$$

The points x, y, z for which equality holds are easy to determine. If $a < 4$, equality occurs in $(**)$ exactly when $x = y = z$ or at least two of x, y, z are 0, and $(*)$ also holds with equality in those cases, so for $a < 4$ the only instances of equality in the triangle given by $x + y + z = 1$ and $x, y, z \geqslant 0$ are the center $(1/3, 1/3, 1/3)$ and the corners $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. When $a = 4$ the situation changes because $(**)$ is not needed, and $(*)$ also gives equality if one of x, y, z is 0 and the other two are equal. Thus if $a = 4$ we also have equality at

$(1/2, 1/2, 0)$, $(1/2, 0, 1/2)$, and $(0, 1/2, 1/2)$, the midpoints of the sides of the above triangle.

Finally, we note that the best inequality of the form $P \geqslant bR$ is $P \geqslant 9R$, which follows from the fact that the entire discussion above applies also when $a = 0$; equality holds in $P \geqslant 9R$ only when $x = y = z$ or when two of x, y, z are zero. The best inequality of the form $P \geqslant aQ$ is $P \geqslant 4Q$, also from the discussion above; equality holds in $P \geqslant 4Q$ only when two of x, y, z are zero or when one of them is zero and the other two are equal.

Solved also by J. Anglesio (France), R. Betts, R. J. Chapman (United Kingdom), E. Cohen (France), R. J. Currier, W. K. Delaney, R. J. Hendel, W. Janous (Austria), I. G. Kastanas, M. E. Kuczma (Poland), L. Kuipers (Switzerland), E. Levine, H. Lipman, K. Liu (West Germany), K. McInturff, H. K. Nickerson, R. Persky, S. Philipp, J. H. Steelman, R. Stong, T. P. Vaughan, K. Zacharias (Germany), and the proposer. Two incorrect solutions were received.

Steiner Trees for the Hypercube

E 3321 [1989, 356]. *Proposed by Robert Spira, Ashland, OR.*

A *Steiner tree* for a finite set S of points in Euclidean n-space is a collection T of straight-line segments such that any two points of S can be joined by a unique path in T; its length is the sum of the segment lengths. It is known that the minimum length of a Steiner tree for the corners of a square of side-length one is $1 + \sqrt{3}$. Show that for any positive integer n there exists a Steiner tree of length $1 + (2^{n-1} - 1)\sqrt{3}$ for the set of vertices of an n-cube of side-length one.

Solution by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. We give a recursive procedure for constructing the desired Steiner tree. The case $n = 1$ is handled by a single line segment.

Assume a Steiner tree of length $1 + (2^{n-1} - 1)\sqrt{3}$ for the set of vertices of an n-cube of side-length one has been constructed, in which each vertex of the n-cube is the endpoint of just one segment and this segment has length greater than $\sqrt{3}/6$. Now consider an $(n + 1)$-cube of side-length one. Select two parallel faces and consider the n-cube whose vertices are the midpoints of the edges joining the corresponding vertices of these faces. Start with the assumed Steiner tree for the vertices of this n-cube. For each vertex a_i of this n-cube, delete a segment of length $\sqrt{3}/6$ from the segment terminating at a_i, then add the segments joining the new endpoint to the two vertices of the $(n + 1)$-cube that are the endpoints of the edge containing a_i. By the Pythagorean Theorem, these segments each have length $\sqrt{3}/3$. The net effect of these changes is to add $2^n(2\sqrt{3}/3 - \sqrt{3}/6) = 2^{n-1}\sqrt{3}$ to the length of the tree. Thus the resulting tree has length $1 + (2^n - 1)\sqrt{3}$. It is clearly a Steiner tree for the vertices of the $(n + 1)$-cube and has the additional properties required for the induction.

Editorial comment. Donnelly Johnson, Marcin Kuczma, and Richard Stong all generalized the result to rectangular parallelopipeds in $\mathbb{R}^n$ of side lengths $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$, obtaining a Steiner tree of length $a_n + \sum_{i=1}^{n-1} 2^{i-1} a_{n-i}\sqrt{3}$.

For $n = 3$ the problem asserts the existence of a Steiner tree of length $1 + 3\sqrt{3} = 6.19615242\ldots$ (for the set of vertices of a three-dimensional cube of side-length one). In the survey paper by E. N. Gilbert and H. O. Pollak, "Steiner

Minimal Trees," *Siam J. Applied Math.*, 16 (1968) 1–29, this numerical value is somewhat garbled.

Solved also by R. J. Chapman (England), D. J. Johnson, M. E. Kuczma (Poland), R. Stong, P. J. Zwier, and the proposer.

An Integral Inequality

E 3322 [1989, 357]. *Proposed by Alexandru Lupaş, Sibiu, Romania.* Suppose g is an even real-valued function on $[-a, a]$, suppose g is nondecreasing on $[0, a]$, and suppose h is a convex real-valued function on $[-a, a]$. Prove that

$$2a\int_{-a}^{a} g(t)h(t)\,dt \geqslant \int_{-a}^{a} g(t)\,dt \int_{-a}^{a} h(t)\,dt.$$

Solution by Duane W. Bailey, Amherst College, Amherst, MA. Let $f(t) = h(t) + h(-t)$ for $0 \leqslant t \leqslant a$. Since g is even, we have $\int_{-a}^{a} g(t)h(t)\,dt = \int_0^a g(t)f(t)\,dt$. Because h is convex, if $0 \leqslant t_1 \leqslant t_2 \leqslant a$, then the chord of the graph of h over the interval $[-t_2, -t_1]$ has no greater slope than the chord over $[t_1, t_2]$, which means that

$$\frac{h(-t_1) - h(-t_2)}{-t_1 - (-t_2)} \leqslant \frac{h(t_2) - h(t_1)}{t_2 - t_1}.$$

This implies that f is nondecreasing. Because both f and g are nondecreasing on $[0, a]$, we can apply Tchebysheff's inequality to conclude

$$\int_0^a dt \int_0^a g(t)f(t)\,dt \geqslant \int_0^a g(t)\,dt \int_0^a f(t)\,dt.$$

Since g is even, translating back to the interval $[-a, a]$ yields the desired inequality.

Editorial comment. Note that convex functions and nondecreasing functions defined on a closed interval are always integrable. Walther Janous observed that the inequality of the problem follows from other inequalities of Lupaş, most published in *Publ. Inst. Math. Univ. Beograd* and quoted in J. E. Pecarić, *Konveksne funkcije. Nejednakosti. Naucna Knjiga.* (Beograd, 1987).

Solved also by U. Abel (West Germany), W. J. Buhler (West Germany), D. Callan, M. Falkowitz (Israel), D. L. Farnsworth, M. Fichter (West Germany), S. K. Ghosh (India), J.-P. Grivaux (France), T. Herman (Hungary), B. Hughes, V. Jianu-Pambuccian, P. G. Kirmser, M. E. Kuczma (Poland), L. Kuipers (Switzerland), K.-W. Lau (Hong Kong), E. Levine, K. Liu (West Germany), L. F. Martins, J.-M. Monier (France), S. Philipp, C. J. Reiss, D. S. Rubin, K. Schilling, R. Stong, K. Zacharias (West Germany), and the proposer.

ADVANCED PROBLEMS

6637. *Proposed by Herbert S. Wilf, University of Pennsylvania, Philadelphia, PA.*

Let $f(n)$ be the sum of the first one-third of the coefficients in the expansion of $(1 + x)^{3n}$, i.e.,

$$f(n) = \sum_{k=0}^{n} \binom{3n}{k} \qquad (n = 0, 1, 2, \ldots).$$

Prove that

$$\sum_{n=0}^{\infty} f(n)\left(\frac{4u^2}{27}\right)^n = \frac{u}{u - 2\sin\left(\frac{1}{3}\arcsin u\right)} - \frac{2u}{2u - 3\sin\left(\frac{1}{3}\arcsin u\right)}.$$

6638. *Proposed by Stan Philipp, Pennsylvania State University, Altoona, PA.*

Let

$$\alpha_k = (-1)^k\binom{-1/2}{k} = 4^{-k}\binom{2k}{k} \qquad (k = 0,1,2,\ldots).$$

(i) Prove that

$$\sum_{k=0}^{\infty} \frac{\alpha_k}{2k+1} = \frac{\pi}{2}, \quad \sum_{k=0}^{\infty} \frac{\alpha_k}{2k-2n+1} = 0 \qquad \text{for } n = 1,2,3,\ldots.$$

(ii) Prove that

$$\sum_{k=0}^{\infty} \alpha_k^2\left(\frac{1}{2k+2n+2} + \frac{1}{2k-2n+1}\right) = 0 \qquad \text{for } n = 0,1,2,\ldots.$$

6639. *Proposed by Paul Monsky, Brandeis University, Waltham, MA.*

Let N be a convex centrally symmetric plane n-gon of area 2 divided into n triangles by rays from an interior point P to the vertices. Let $a_1, a_2, \ldots, a_n$ be the areas of these triangles, taken in cyclic order.

(i) If N is affinely equivalent to the regular polygon of $2k$ sides, prove that

$$a_1 + a_3 + \cdots + a_{2k-1} = a_2 + a_4 + \cdots + a_{2k} = 1.$$

(ii) In general prove that $a_1, a_2, \ldots, a_n$ satisfy a polynomial identity with integer coefficients and constant term 1.

SOLUTIONS OF ADVANCED PROBLEMS

Squares Expressible As a Sum of n Consecutive Squares

6552 [1987, 694]. *Proposed by Moshe Laub, Jerusalem, Israel.*

Let S be the set of positive integers n for which there exists a square equal to the sum of the squares of n consecutive positive integers. For example, $11 \in S$, since

$$77^2 = 18^2 + 19^2 + \cdots + 27^2 + 28^2$$

Prove that

(a) S has density zero,

(b) S is infinite,

(c) if $n \in S$ and n is not itself a square, then there are infinitely many squares each of which is equal to the sum of the squares of n consecutive positive integers.

Further, since $x_0 \geqslant -(n-1)/2$ and $y_0 > (n-1)/2$, we have $y \geqslant y_0 u > 0$ and

$$x \geqslant y_0 v - u(n-1)/2 + (n-1)(u-1)/2 = y_0 v - (n-1)/2$$
$$\geqslant y_0 - (n-1)/2 > 0.$$

Thus the theorem is proved.

Editorial comment. If $N(x)$ is the number of elements of S not exceeding the positive number x, the argument of either solution of (a) shows that in fact $N(x) = O(x(\log x)^{-1/2})$. (Cf. §183 of Landau's *Handbuch der Lehre von der Verteilung der Primzahlen*.) Using the arguments of the two solutions of (a) along with upper-bound sieve methods, Heini Halberstam and Hans-Egon Richert provided us with a proof that actually $N(x) = O(x(\log x)^{-1})$. In the other direction either solution of (b) shows that $N(x) > c\sqrt{x}$ for a suitable positive constant c. It seems difficult to determine the precise order of magnitude of $N(x)$.

The elements of S less than 73, along with corresponding values of x for which $T_n(x)$ is a perfect square, are given in the following table

n	1	2	11	23	24	26	33	47	49	50	59
x	1	3	18	7	1	25	7	539	25	7	22

Most of these examples were given by E. Barbette in a monograph published in 1910. Cf. L. E. Dickson, *History of the Theory of Numbers*, Volume II, pp. 319–323.

A problem distantly related to the present one is E 3064 [1984, 649; 1987, 190].

The solutions by Lossers, Mattics, and the editors were the only ones received.

Infinitely Differentiable Functions on $\mathbb{R}^7$

6579 [1988, 773–774]. *Proposed by D. K. Cohoon, Temple University, Philadelphia, PA.*

Suppose we define a curl operation on the infinitely differentiable functions from $\mathbb{R}^7$ to $\mathbb{C}^7$ by the rule

$$\operatorname{curl}(\vec{f}) \doteq \sum_{i=1}^{7} \left\{ \left(\frac{\partial f_{i+3}}{\partial x_{i+1}} - \frac{\partial f_{i+1}}{\partial x_{i+3}} \right) + \left(\frac{\partial f_{i+6}}{\partial x_{i+2}} - \frac{\partial f_{i+2}}{\partial x_{i+6}} \right) + \left(\frac{\partial f_{i+5}}{\partial x_{i+4}} - \frac{\partial f_{i+4}}{\partial x_{i+5}} \right) \right\} \vec{e}_i$$

where $\vec{e}_i$ is the unit vector in the positive direction on the ith coordinate axis, $\vec{f} = \vec{f}(x) = f_1\vec{e}_1 + f_2\vec{e}_2 + f_3\vec{e}_3 + f_4\vec{e}_4 + f_5\vec{e}_5 + f_6\vec{e}_6 + f_7\vec{e}_7$ is a C^∞ function from $\mathbb{R}^7$ to $\mathbb{C}^7$, and, for all positive integers j in $\{1,2,3,4,5,6,7\}$, $x_{j+7} = x_j$ and $f_{j+7} = f_j$. Show that every infinitely differentiable function from $\mathbb{R}^7$ to $\mathbb{C}^7$ is a curl plus a gradient.

Solution by Alfred Gray, Department of Mathematics, University of Maryland, College Park, Maryland. The operational identity

$$\operatorname{curl}(\operatorname{curl}) = \operatorname{grad}(\operatorname{div}) - \Delta$$

can be verified for any infinitely differentiable vector field by a slightly lengthy but purely mechanical process. One simply checks that the ith component on each side is the same (details omitted). Since the Laplacian is an epimorphism on the infinitely differentiable functions on every open set, the result follows.

Solved also by the proposer.

The Asymptotic Expansion of the Middle Binomial Coefficient

6581 [1988, 774]. *Proposed by Donald E. Knuth and Ilan Vardi, Stanford University, Stanford, CA.*

Stirling's approximation leads to the following expansion

$$\binom{2n}{n} = \frac{4^n}{\sqrt{\pi n}}\left\{1 - \frac{1}{8n} + \frac{1}{128n^2} + \frac{5}{1024n^3} - \frac{21}{32768n^4} + O\left(\frac{1}{n^5}\right)\right\}.$$

Prove that if this expansion is continued indefinitely, the coefficient of n^{-k} will be $2^{1-4k}m_k$, where m_k is an integer. What is the exact power of 2 that divides the denominator of this coefficient?

Solution I *by Rolf Richberg, Institut für reine und angewandte Mathematik der RWTH Aachen, West Germany.* For $x > 0$, define

$$f(x) = \sqrt{x}\,\frac{\Gamma\left(x + \frac{1}{2}\right)}{\Gamma(x+1)} = \frac{\sqrt{\pi x}}{4^x}\,\frac{\Gamma(2x+1)}{\Gamma(x+1)^2}, \tag{1}$$

where the equality of the two expressions follows from the duplication formula for the gamma function. By Stirling's formula (or formula 6.1.47 on p. 257 of Abramowitz and Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, 1964) we have an asymptotic expansion

$$f(x) \sim \sum_{m=0}^{\infty} a_m x^{-m} \tag{2}$$

with $a_0 = 1$. Since $f(n) = \sqrt{\pi n}\binom{2n}{n}/4^n$, the problem is to describe the coefficients a_m in (2). Denote by $\alpha(m)$ the number of ones in the binary expansion of $m \in \mathbb{N}$, where $\mathbb{N}$ denotes the positive integers. We shall show that for any $m \in \mathbb{N}$,

$$a_m = 2^{\alpha(m)-4m}\,a_m^*, \tag{*}$$

where a_m^* is an odd integer.

For the proof of (*) it is convenient to put

$$b_0 = 1, \quad b_m = (-1)^m\,2^{4m-1}\,a_m, \qquad m \in \mathbb{N}, \tag{3}$$

and to verify (*) in the equivalent form

$$b_m = 2^{\alpha(m)-1}\,b_m^*, \qquad m \in \mathbb{N}, \tag{**}$$

where b_m^* is an odd integer. To do this, we shall set up two functional equations for f, each of which gives a recurrence relation for the b_m. The first will show that b_m is always an integer, while the second will yield (**).

From (1) we have

$$f(x)f\left(x + \frac{1}{2}\right) = \left(1 + \frac{1}{2x}\right)^{-1/2} \qquad x > 0. \tag{4}$$

From (2) and the binomial expansion

$$\left(x+\frac{1}{2}\right)^{-k} = x^{-k}\sum_{j=0}^{\infty}\binom{-k}{j}(2x)^{-j}$$

we derive the asymptotic expansion

$$\begin{aligned} f\left(x+\frac{1}{2}\right) &\sim \sum_{k=0}^{\infty} a_k x^{-k}\sum_{j=0}^{\infty}\binom{-k}{j}2^{-j}x^{-j} \\ &= \sum_{m=0}^{\infty}\left(\sum_{k=0}^{m}\binom{-k}{m-k}2^{k-m}a_k\right)x^{-m} \qquad (x\to\infty). \end{aligned} \tag{5}$$

Upon multiplying the expansions (2) and (5), we obtain from (4) the expansion

$$\sum_{m=0}^{\infty}\binom{-\frac{1}{2}}{m}2^{-m}x^{-m} = \sum_{m=0}^{\infty}\left(\sum_{k=0}^{m}a_{m-k}\sum_{j=0}^{k}\binom{-j}{k-j}2^{j-k}a_j\right)x^{-m} \qquad (x\to\infty).$$

By equating coefficients and using the identity

$$\binom{-\frac{1}{2}}{m} = \frac{(-1)^m}{4^m}\binom{2m}{m}$$

we obtain

$$\begin{aligned} (-1)^m 2^{-3m}\binom{2m}{m} &= \sum_{k=0}^{m}\sum_{j=0}^{m-k}\binom{-j}{m-k-j}2^{j+k-m}a_j a_k \\ &= \sum_{k=0}^{m}\sum_{j=0}^{m-k}\binom{m-k-1}{m-k-j}\left(-\frac{1}{2}\right)^{m-k-j}a_j a_k. \end{aligned}$$

Since $a_0 = 1$ and $a_m = (-1)^m b_m 2^{1-4m}$ for $m \in \mathbb{N}$, this is easily rearranged to give

$$\begin{aligned} b_m &= 2^{m-2}\binom{2m}{m} - \sum_{j=1}^{m-1}\binom{m-1}{j-1}2^{3(m-1-j)+2}b_j \\ &\quad - \sum_{k=1}^{m-1}\sum_{j=1}^{m-k}\binom{m-k-1}{j-1}2^{3(m-k-j)}b_j b_k \end{aligned}$$

for $m \in \mathbb{N}$; this shows that $b_1 = 1$ and that (by induction on m)

$$b_m \text{ is an integer for } m \in \mathbb{N}. \tag{6}$$

The second functional equation for f, namely

$$f(x+1) = \left(1+\frac{1}{2x}\right)\left(1+\frac{1}{x}\right)^{-1/2} f(x), \qquad x > 0, \tag{7}$$

is also an immediate consequence of (1). The binomial expansion

$$(x+1)^{-k} = x^{-k}\sum_{j=0}^{\infty}\binom{-k}{j}x^{-j},$$

together with (2), yields

$$f(x+1) \sim \sum_{m=0}^{\infty} \left(\sum_{k=0}^{m} \binom{-k}{m-k} a_k \right) x^{-m} \qquad (x \to \infty).$$

Upon multiplying the power series expansion of the first two factors on the right-hand side of (7) with the asymptotic expansion (2), we find that

$$\sum_{m=0}^{\infty} \left[\sum_{k=0}^{m} \binom{-k}{m-k} a_k \right] x^{-m}$$

$$= \sum_{m=0}^{\infty} \left\{ a_m + \sum_{k=0}^{m-1} \left[\binom{-\frac{1}{2}}{m-k} + \frac{1}{2} \binom{-\frac{1}{2}}{m-k-1} \right] a_k \right\} x^{-m} \qquad (x \to \infty).$$

Now equate coefficients to obtain (for $m \in \mathbb{N}$)

$$0 = \sum_{k=0}^{m-1} \left[\binom{-k}{m-k} - \binom{-\frac{1}{2}}{m-k} - \frac{1}{2} \binom{-\frac{1}{2}}{m-k-1} \right] a_k$$

$$= \sum_{k=1}^{m} \left[(-1)^k \binom{m-1}{k} - (-1)^k 2^{1-2k} \frac{k-1}{k} \binom{2k-2}{k-1} \right] a_{m-k}.$$

In view of (3) this may be rewritten (by shifting the summation on k and putting $m+1$ in place of m) as

$$m b_m = 2^{2m-2} \frac{m}{m+1} \binom{2m}{m} + \sum_{k=1}^{m-1} \left[2^{2k-1} \frac{k}{k+1} \binom{2k}{k} - 2^{4k} \binom{m}{k+1} \right] b_{m-k} \quad (8)$$

for $m \in \mathbb{N}$.

Now define $\beta(m)$ for any $m \in \mathbb{N}$ by $m = 2^{\beta(m)} m^*$, where m^* is an odd integer. Let $2^{\delta(m)}$ be the lowest power of 2 *not* occurring in the binary expansion of m. Then $\delta(m) = \beta(m+1)$ and we may write

$$m = 2^0 + 2^1 + \cdots + 2^{\delta-1} + 2^{s_\delta} + 2^{s_{\delta+1}} + \cdots + 2^{s_{\alpha-1}}$$

with $0 \leqslant \delta < s_\delta < s_{\delta+1} < \cdots < s_{\alpha-1}$, where $\alpha = \alpha(m)$ and $\delta = \delta(m)$. It follows that $m+1 = 2^\delta + 2^{s_\delta} + 2^{s_{\delta+1}} + \cdots + 2^{s_{\alpha-1}}$. Hence

$$\alpha(m+1) = \alpha(m) + 1 - \beta(m+1), \qquad m \in \mathbb{N}, \quad (9)$$

and consequently

$$m - \alpha(m) \leqslant m + 1 - \alpha(m+1), \qquad m \in \mathbb{N}. \quad (10)$$

We now prove (**) by induction on m. Since $b_1 = 1$ and $\alpha(1) = 1$, (**) is trivially satisfied for $m = 1$, and we may assume that for some $m \geqslant 2$,

$$b_{m-k} = 2^{\alpha(m-k)-1} b^*_{m-k}, \qquad k = 1, 2, \ldots, m-1,$$

where b^*_{m-k} is an odd integer. Then (8) becomes

$$mb_m = 2^{2m-2}\frac{m}{m+1}\binom{2m}{m} + \left\{1 - 8\binom{m}{2}\right\}2^{\alpha(m-1)}b^*_{m-1} \tag{11}$$
$$+ \sum_{k=2}^{m-1}\left[2^{k-2}\frac{k}{k+1}\binom{2k}{k} - 2^{3k-1}\binom{m}{k+1}\right]2^{\alpha(m-k)+k}b^*_{m-k}.$$

From (10) we find that $\alpha(m-k)+k \geqslant \alpha(m-1)+1$ for $k = 2, \ldots, m-1$, and we have, trivially,

$$2m - 2 \geqslant \alpha(m-1) + m - 1 \geqslant \alpha(m-1) + 1.$$

Thus, all terms on the right-hand side of (11) are integer multiples of $2^{\alpha(m-1)+1}$, with the sole exception of $2^{\alpha(m-1)}b^*_{m-1}$. This shows that the number

$$2^{-\alpha(m-1)}mb_m = 2^{\beta(m)-\alpha(m-1)}m^*b_m = 2^{1-\alpha(m)}m^*b_m$$

(recall (9)) is an odd integer. Since m^* is odd and b_m is an integer by (6), this concludes the proof of (**).

Solution II *by Donald E. Knuth, Stanford University, Stanford, California.* By the duplication formula for the gamma function, the given formula can be rewritten as

$$\frac{\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n)n^{1/2}} \sim 1 - \frac{1}{8n} + \cdots.$$

We shall prove that if p/q is any rational number,

$$\frac{\Gamma(n+p/q)}{\Gamma(n)n^{p/q}} \sim 1 + \frac{m_1(p,q)}{q^3n} + \frac{m_2(p,q)}{q^7n^2} + \frac{m_3(p,q)}{q^{11}n^3} + \cdots,$$

where each $m_k(p,q)$ is an integer. Tricomi and Erdélyi (*Pacific J. Math.*, 1(1951) 133–142) showed that such an expansion exists with

$$m_k(p,q)/q^{4k-1} = M_k(p/q)$$

where $M_k(t)$ is a polynomial of degree $2k$. This polynomial can be identified as a Stirling number of the first kind, because

$$\frac{\Gamma(n+t)}{\Gamma(n)} = n(n+1)\cdots(n+t-1)$$

when t is a positive integer. The polynomial is well known to be a linear combination of binomial coefficients,

$$M_k(t) = \sum_{j=1}^{k} C_{k,j}\binom{t+j-1}{2k},$$

where the $C_{k,j}$ are integers (see Jekuthiel Ginsburg, "Note on Stirling's numbers," this MONTHLY 35(1928), 77–80; Ira Gessel and Richard P. Stanley, "Stirling polynomials," *J. Comb. Theory* (A) 24(1978), 24–33). Therefore the result follows from Hermite's theorem that

$$\binom{p/q}{k}q^{2k-1}$$

is an integer (*Cours de Hermite*, Faculté de Science de Paris, 1882, Leçon 19).

The result can be sharpened when $q = 2$, because $C_{k,1} = 1$ and $C_{k,j}$ is even for $j > 1$. Therefore the exact power of 2 dividing the denominator of the coefficient of n^{-k} is $4k - \nu(k)$, where $\nu(k)$ is the number of ones in the binary representation of k.

Editorial remark. A problem essentially identical to this one was proposed to the MONTHLY in 1985 by Douglas B. Tyler, then at the University of California at Davis. There was, however, no proof of the formula involving the number of ones in the binary expansion of n, and the problem was rejected at that time. In 1986 Dean Hickerson responded to an inquiry from Tyler with a ten page manuscript about recurrences and Stirling numbers that included a solution of Tyler's problem. In 1988 the current advanced problem editors (an entirely different set of people) received and accepted the Knuth-Vardi problem proposal. As indicated above, this proposal *was* accompanied by a proof of the binary expansion formula. Upon its publication another Hickerson-Tyler correspondence ensued. In particular, a November 27, 1988 letter of Hickerson established that the full solution of the Knuth-Vardi problem was an immediate consequence of his 1986 manuscript. The editors have examined Hickerson's letters and the letter rejecting Tyler's original proposal, and publish this note to acknowledge the priority of Tyler and Hickerson.

For other relations between binomial coefficients and expansions of integers in various bases (including base 2) see N. J. Fine, "Binomial coefficients modulo a prime," this MONTHLY 54(1947), 589–592, and also the forthcoming book (in Russian) "Generalized Pascal's Triangles and Pyramids, their Fractals, Graphs, and Applications" by the Soviet mathematician B. A. Bondarenko (Tashkent).

The solutions by Hickerson, Knuth, and Richberg were the only ones received.

A Condition Implying Nilpotence

6584 [1988, 880]. *Proposed by E. T. Parker, University of Illinois at Urbana-Champaign.*

Suppose a finite group G satisfies the following condition: If H is a cyclic subgroup of G of prime-power order or if H is an elementary abelian subgroup of G, then H lies in the center of the normalizer of H in G.

(a) Show that G need not be abelian.
(b) Show that G must be nilpotent.

Solution by Richard M. Stafford, Department of Defense, Fort George Meade, Maryland. For (a), the Suzuki 2-groups of type $A(n, \theta)$ provide an infinite family of examples. Such a group can be realized as the group of upper triangular matrices of the form

$$\begin{pmatrix} 1 & a & b \\ 0 & 1 & a^{\theta} \\ 0 & 0 & 1 \end{pmatrix},$$

where a and b are members of the finite field $GF(2^n)$ and θ is a nontrivial field automorphism of odd order.

These groups are of exponent 4 with the property that all involutions lie in the center. Thus we need only check that H lies in the center of the normalizer of H for cyclic subgroups H of order 4. But this fact is an immediate consequence of Lemma 11.3 (p. 265) of B. Huppert and N. Blackburn's *Finite Groups III*, Springer-Verlag, Berlin-Heidelberg-New York, 1982.

We now solve (b). For brevity we refer to groups satisfying the condition of the problem as Z-groups. The rest of our notation as well as all results we need for our solution may be found in D. Gorenstein's *Finite Groups*, Harper and Row, New York, 1968. We shall refer to this reference as **[GOR]**. We now need a lemma.

LEMMA. *Let G be a Z-group. Then the following hold*:

(i) *Every subgroup of G is a Z-group.*

(ii) *If G is a p-group, then $\Omega_1(G)$ (the subgroup generated by the elements of order p) lies in the center of G. In particular $\Omega_1(G)$ is elementary abelian.*

Proof. Part (i) is obvious. So assume G is a p-group and let E be a maximal elementary abelian p-subgroup of G. Since G is a Z-group, it suffices to show that E is a normal subgroup of G. Set $N = N_G(E)$ and assume by way of contradiction that N is a proper subgroup of G. Then N is a proper subgroup of $N_G(N)$ because G is a p-group (Theorem 1.2.11 (ii) of **[GOR]**). Now, by the maximality of E and the fact that $E \subseteq Z(N)$, it follows that $E = \Omega_1(N)$. Thus E is a characteristic subgroup of N. But this implies $N_G(N) = N$ and we have a contradiction.

We are now ready to prove that Z-groups are nilpotent. It suffices to prove G has a normal p-complement for each prime p which divides the order of G. By the Frobenius normal p-complement theorem (Theorem 7.4.5 of **[GOR]**) this is equivalent to showing that the factor group of $N_G(H)$ with respect to $C_G(H)$ is a p-group for every nonidentity p-subgroup H of G. In the remainder of the proof let $N = N_G(H)$ and $C = C_G(H)$ for a particular nonidentity p-subgroup H of G. Then, by our lemma and the fact that G and H are Z-groups, it follows that $N \subseteq C_G(\Omega_1(H))$. Now each element of $N - C$ induces by conjugation an automorphism of H. The fact that N/C is a p-group follows quickly from the fact that any automorphism of H whose order is relatively prime to p must act nontrivially on $\Omega_1(H)$ provided p is odd (see Theorem 5.3.10 of **[GOR]** for details). So we assume p is 2. We now proceed by contradiction and assume H (our 2-subgroup of G) is of minimal order subject to the order of N/C being divisible by an odd prime. Let $y \in N - C$ be of odd order. It follows from the minimality of H that, if y normalizes a proper subgroup of H, then y centralizes that subgroup. The structure of H is fairly well determined by Theorem 5.3.7 of **[GOR]**. In particular H is either elementary abelian or H is of class 2, $Z(H)$ is elementary abelian and $H/Z(H)$ elementary abelian. If H is elementary abelian, then $y \in C$ as G is a Z-group and we have a contradiction. So we may assume the latter case holds. In particular H has exponent 4 and y centralizes $Z(H)$. The fact that $Z(H)$ is elementary abelian together with our lemma implies $\Omega_1(H) = Z(H)$. We are now ready to obtain a contradiction. Since $y \in N - C$ there exist distinct elements of order 4, u and v, with $y^{-1}uy = v$. Moreover since y has odd order it follows that u and v lie in distinct cosets of $Z(H)$ in H. Thus uv has order 4 also. Now $[u, v]$ commutes with u and v because H is a class 2 group. Thus by Lemma 2.2.2 of **[GOR]**, $(uv)^2 = [v, u]u^2v^2$. But u^2 and $v^2 \in \Omega_1(H)$ because H is of exponent 4. Thus $v^2 = (y^{-1}uy)^2 = y^{-1}u^2y = u^2$ and we have $(uv)^2 = [v, u]$. Since uv is of

order 4, $(uv)^{-1} = (uv)(uv)^2 = uv[v, u] = vu$. So $u^{-1}(uv)u = vu = (uv)^{-1}$. Thus u normalizes the cyclic group $\langle uv \rangle$ of order 4. Since uv has order 4, u does not centralize uv. This contradicts the fact that G is a Z-group.

Editorial comment The condition $N(H) = C(H)$ when H is a cyclic p-group having order greater than p is used only when $p = 2$ and $|H| = 4$. Thus the conclusion of the problem remains valid if we assume $N(H) = C(H)$ only when H is an elementary abelian subgroup or when H is a cyclic subgroup of order 4.

A similar solution was provided by Gary L. Walls. The editors wish to thank Everett C. Dade and Stephen M. Gagola, Jr., for their help with this problem.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Lattice Points. By Paul Erdös, Peter M. Gruber and Joseph Hammer, Pitman Monographs and Surveys in Pure Mathematics 39, Longman Scientific and Technical, co-published with John Wiley and Sons, New York, 1989, vii + 184 pp.

WILLIAM FIREY
Department of Mathematics, Oregon State University, Corvallis, OR 97331

Pythagoras' theorem and Euclid's proof that there are infinitely many prime numbers are initial landmarks for geometry and elementary number theory. They show the great age of these parts of mathematics. So it is surprising that fruitful interactions between these disciplines came, almost all, within the last 100 years, more than two millenia after Euclid. To illustrate the interaction I choose first one of the earliest results of this sort: Minkowski's theorem about lattice points in convex sets, published in 1893 as a paper read at the International Mathematical Congress at Chicago. It asserts that a compact convex set C in $\mathbb{R}^d$ which is symmetric with respect to the origin o and has volume at least 2^d must contain a point, other than o, with integer coordinates. The idea of the proof is as simple and elegant as the theorem's statement. Suppose $C \setminus \{o\}$ has no points with integer coordinates. Form all integer translates of $(\frac{1}{2})C$, that is sets of the form

$$(\tfrac{1}{2})C + n = \{x = (x_1, \ldots, x_d) : x = (\tfrac{1}{2})y + n, \quad y = (y_1, \ldots, y_d) \in C, \quad n = (n_1, \ldots, n_d) \in \mathbb{Z}^d\}.$$

Use the convexity of C and the assertion that $C \setminus \{o\}$ has no points of $\mathbb{Z}^d$ to prove that the sets $(\frac{1}{2})C + n$ are disjoint. Also these sets meet any unit cube, cornered at points of $\mathbb{Z}^d$, in pieces of total volume equal to that of $(\frac{1}{2})C$. This covered volume is $(\frac{1}{2})^d$ times the volume of C; it is also less than the volume of the unit cube. Hence C has volume less than 2^d.

The theorem begins what Minkowski called the geometry of numbers. It has beautiful applications. Here is one. If $A = [a_{ij}]$ is a symmetric, positive definite $d \times d$ matrix, then, for $\alpha > 0$,

$$\left\{x : Q(x) = \sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij} x_i x_j \leqslant \alpha\right\}$$

is an ellipsoidal solid, centered at o, of volume $\alpha^{d/2}\kappa_d/(\det A)^{1/2}$, where κ_d is the volume of the Euclidean unit ball in $\mathbb{R}^d$. In particular, with $\alpha = 4(\det A/\kappa_d^2)^{1/d}$, this ellipsoid has volume 2^d. Minkowski's theorem shows that

$$\min_{\mathbb{Z}^d \setminus \{o\}} Q(x) \leqslant 4\left(\det A/\kappa_d^2\right)^{1/d}$$

Also the same theorem gives information about solutions to Diophantine problems of the form: find n in $\mathbb{Z}^d$ such that, with $\langle v, n\rangle = \sum_{i=1}^{d} v_i n_i$

$$|\langle u_1, n\rangle| \leqslant \tau_1, \ldots, |\langle u_d, n\rangle| \leqslant \tau_d$$

for given positive real numbers $\tau_1, \ldots, \tau_d$ and vectors $u_1, \ldots, u_d$ which are independent, so that the determinant $\det[u_1, \ldots, u_d]$ having these vectors as columns is not zero. Take C to be the convex, compact polyhedral solid of points $\pm x$ which satisfy

$$|\langle u_1, x\rangle| \leqslant \tau_1, \ldots, |\langle u_d, x\rangle| \leqslant \tau_d;$$

its volume is $2^d \tau_1 \cdots \tau_d / |\det[u_1 \cdots u_d]|$. Minkowski's theorem shows that there are solutions if $\tau_1 \cdots \tau_d \geqslant |\det[u_1 \cdots u_d]|$.

Notice how Minkowski's theorem and its applications use interplays of quite simple notions from elementary number theory and geometry. To appreciate the power of this interplay the reader might attempt purely number-theoretic treatments of the first application with $d = 2$. Assistance can be found in reference [1, pp. 30–31].

Notice also the importance of the set $\mathbb{Z}^d$, the simplest geometric lattice in $\mathbb{R}^d$. More generally, a lattice in $\mathbb{R}^d$ is the set of all linear combinations with integer coefficients of vectors which form a basis for $\mathbb{R}^d$. The book under review surveys those topics in which geometric lattices play an important role, the geometry of numbers being but one such.

The work is like a Baedeker's travel guidebook: all the sights are listed, with extensive historical backgrounds for the important ones—in a few major cases there are detailed descriptions; excellent maps are provided. A typical chapter surveys a topic with a summary of its history, a discursive description of some major results, and closes with a nicely classified summary of recent advances. All three authors are noted experts in the areas covered; they are clear expositors and meticulous scholars; they have produced a fine book.

The extent of the country covered is large. Here are located parts of two of Hilbert's 23 problems presented to the 1900 International Congress of Mathematicians at Paris; problem 3 on the equidissectability of polyhedra and problem 18 in which Hilbert included the questions of densest packings of equal solids in space. Going out from Pick's elegant formula for the area enclosed by a closed Jordan polygon with vertices in $\mathbb{Z}^2$, the authors lead us to the algebraic-geometric notion of a toric variety and its connection with new proofs of the Fenchel-Aleksandrov inequalities for convex compacta. Hilbert's 18th problem is met again in the study of tiling problems and crystallographic groups.

There is an excursion through positive quadratic forms, beginning with an equivalence notion: Q, Q' are equivalent precisely when $Q'(x) = Q(Tx)$ for some $d \times d$ integer matrix T with determinant ± 1. This connects with lattices. Generate a lattice L from a basis $\{b_1, \ldots, b_d\}$; form the matrix B having these vectors as its columns; then

$$Q(x) = x^{\mathrm{tr}} B^{\mathrm{tr}} B x$$

is a positive quadratic form associated with L. Q determines L up to a rotation—possibly improper. A different basis for L produces another metric form Q'; Q, Q' will be equivalent. Thus the study of possible bases for L amounts to the

study of forms Q' equivalent to Q. This excursion ends with a survey of locally most dense lattice packings of equal balls. For each lattice L there is a maximum number $\rho(L)$ such that packed balls, centered at points of L, have disjoint interiors. As before, let $B = B(L)$ signify the matrix associated with a basis for L. $|\det B(L)|$ is the volume of the parallelohedral solid or cell determined by the basis for L. The total volume of the pieces of packed balls falling within one cell equals the volume of a ball of radius $\rho(L)$. Hence,

$$r(L) = (\rho(L))^d \kappa_d / |\det B(L))|$$

measures the packing density. It is locally most dense just when $r(L) \geqslant r(L')$ for all lattices L' near enough to L in a suitable sense.

On the frontiers can be found power diagrams, zeta-functions for lattices, dimers, rook polynomials, Steiner trees, polyominoes—also known as n-celled animals—and other exotica.

A good guidebook will lure the armchair traveler into an actual visit: this is a good guidebook.

REFERENCE

1. J. W. S. Cassels, An Introduction to the Geometry of Numbers, Die Grundlehren der mathematischen Wissenschaften, vol. 99, Springer-Verlag, Berlin, Göttingen, Heidelberg, 1959.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S*, P*, L.** *Islands of Truth: A Mathematical Mystery Cruise.* Ivars Peterson. WH Freeman, 1990, xvii + 325 pp, $19.95. [ISBN: 0-7167-2113-9] "The pursuit of mathematical truth is a tricky business, full of mysteries and surprises." This vivid, compelling, pleasure-filled guide reveals to the mathematical tourist many mysteries of modern mathematics: knots, tilings, computers, fractals, dimensions, surfaces, chaos, decompositions, NP-problems, and much, much more. Based on eight years of reporting on mathematics for *Science News*, both the choice of and approach to topics provide ample evidence of the reporter's craft at its very best. An excellent, innovative option for any mathematics course used for general education. LAS

General, P*, L. *Renewing U.S. Mathematics: A Plan for the 1990s.* National Research Council. National Academy Pr, 1990, xiv + 133 pp, (P). [ISBN: 0-309-04228-3] This is "David II," the sequel to the 1984 ground-breaking report from a committee of the National Academy of Sciences chaired by Edward E. David, Jr. This report recommends continued effort to achieve the goals of the 1984 document ("stay the course; see it through") modified, however, by a special focus on human resource issues. Three primary recommendations address increasing federal funding for mathematical research to $250 million per year; improving career paths in the mathematical sciences; and increasing to 2600 the number of supported senior investigators. Special feature: a 45-page appendix giving brief descriptions of 27 different areas of research accomplishments and related opportunities. LAS

Precalculus, T(13: 1, 2). *Precalculus: Functions and Graphs.* Linda Gilbert, Jimmie Gilbert. Addison-Wesley, 1990, xiv + 615 pp, $32.76. [ISBN: 0-201-51020-0] A pre-precalculus text covering standard range of topics. Many routine problems, plus a number of problems resembling calculus I min/max applications. Each chapter has 1-3 pages of "non-solutions" to problems, enabling students to see typical errors. FR

Finite Mathematics, T(14-15: 1). *Applied Finite Mathematics, Third Edition.* S.T. Tan. PWS/Kent, 1990, xiv + 543 pp, $35. [ISBN: 0-534-92140-X] Changes from the *Second Edition* are relatively minor and technical. (*First Edition,* TR, June-July 1983; *Second Edition,* TR, June-July 1987.) JS

Foundations, T(18), P, L. *Lecture Notes in Mathematics-1407: Proof Theory, An Introduction.* Wolfram Pohlers. Springer-Verlag, 1989, vi + 213 pp, $20 (P). [ISBN: 0-387-51842-8] A self-contained introduction to pure number theory, to the axiomatization of the natural numbers (with careful consideration for consistency of the axioms), and to ordinal analysis, the computation of the least ordinal not provable in a formal theory. Follows Hilbert's program which can be roughly characterized as an attempt to axiomatize the whole of mathematics and to prove that the axioms are consistent. SB

Number Theory, T*(16-18), S, P**, L*.** *Algorithmic Algebraic Number Theory.* M. Pohst, H. Zassenhaus. Encyclopedia of Math. & Its Applic. Cambridge U Pr, 1989, xiv + 465 pp, $89.50. [ISBN: 0-521-33060-2] Introduction to constructive algebraic number theory and some associated algorithms, assuming a basic algebra background. Covers the determination of the Galois group of an equation in $\mathbf{Q}[x]$, the computation of the maximal order of the unit group of an algebraic number field, and the ideal class group. Appendix includes extensive numerical tables and a list of algorithms presented in the text. A nice companion text (with exercises) for a first course in algebraic number theory. Great reference for expert and non-expert alike. SB

Number Theory, P. *Lecture Notes in Mathematics-1415: Cinquante Ans de Polynômes: Fifty Years of Polynomials.* Eds: M. Langevin, M. Waldschmidt. Springer-Verlag, 1990, ix + 235 pp, $21.80 (P). [ISBN: 0-387-52190-9] Proceedings of a conference held in honor of Alain Durand (1949-86) at the Institut Henri Poincaré in Paris, May 26-27, 1988. Eleven papers from the conference, plus four by Durand. All but one are in French. DFA

Linear Algebra, T(16-17: 1), S, L. *Lectures on*

Linear Algebra. I.M. Gel'fand. Dover, 1989, 185 pp, $5.95 (P). [ISBN: 0-486-66082-6] Concise treatment of basic ideas of linear algebra at a relatively advanced level. Not an introductory-level text. Chapters on n–dimensional spaces, linear and bilinear forms; linear transformations; canonical forms; introduction to tensors. LC

Algebra, S(18), P. *Clifford Theory for Group Representations.* Gregory Karpilovsky. Math. Stud., V. 156. North-Holland (US Distr: Elsevier Science), 1989, x + 364 pp, $110.50. [ISBN: 0-444-87377-5] Intended to present state of developments originating from Clifford's work begun in 1937 to obtain irreducible representations for a finite group G from those for a normal subgroup N. Assumes a graduate course in algebra. A survey chapter is followed by chapters on restriction, induction, and extension. Bibliography, index. JS

Algebra, S(18), P. *Unit Groups of Group Rings.* Gregory Karpilovsky. Pitman Mono. & Surveys in Pure & Appl. Math., V. 47. Longman Scientific & Technical (US Distr: Wiley), 1989, xiii + 393 pp, $143. [ISBN: 0-582-02341-6] Assuming a graduate course in algebra as background, first three chapters discuss general group rings and the group of units. Subsequent chapters deal with unit groups for commutative groups, work of Roggenkamp and Scott on conjugacy, and metabelian integral group rings. Bibliography, index. JS

Calculus, T(13). *Calculus, An Integrated Approach.* Donald B. Small, John M. Hosack. Intern. Ser. in Pure & Appl. Math. McGraw-Hill, 1990, xviii + 685 pp, $34.95. [ISBN: 0-07-058264-5] Here is a calculus text that is truly new. It is a new book, not a new edition, from a new author, written in a new way. Familar in the MAA as a pioneer in computer algebra systems, Don Small has given us a lean book that should be examined by everyone who is sympathetic to the spirit of calculus reform. AWR

Calculus, T*(13: 1, 2). *Calculus for the Managerial, Life, and Social Sciences, Second Edition.* S.T. Tan. PWS-Kent, 1990, xiv + 568 pp, $35. [ISBN: 0-534-92145-0] For a one-semester or two-quarter introductory course. Intuitive; informal; ample examples, exercises, problems (odd answers); numerous figures, nicely done and helpful. Partial solutions manual available to students; answer book, test book, computerized test bank available to adopters; graphics/computation software package accompanies book but was not seen by reviewer. Changes in this edition (*First Edition,* TR, May 1988) include preliminary chapter on precalculus; boxed definitions, methods and procedures; so-labelled calculator exercises; earlier introduction of integration. JDEK

Real Analysis, T(15). *Advanced Calculus, Second Edition.* David V. Widder. Dover, 1989, xvi + 520 pp, $10.95 (P). [ISBN: 0-486-66103-2] A classic text written when advanced calculus took up where calculus left off and pushed forward in areas felt to be important to applied mathematicians and engineers. It was never doubted, of course, that pure mathematicians also needed to cover this same material, but their ultimate concerns (proofs, examination of deeper concepts) were relegated to the last part of each chapter. AWR

Complex Analysis, S(18), P. *Several Complex Variables I: Introduction to Complex Analysis.* Ed: A.G. Vitushkin. Encyclop. of Math. Sci., V. 7. Springer-Verlag, 1990, 248 pp, $59. [ISBN: 0-387-17004-9] Five long survey articles, in expository style, on fundamental topics in the theory of functions of several complex variables: basic ideas, integral representations, analytic varieties, holomorphic mapping, residue theory. With their extensive bibliographies, the articles provide a wide overview of modern work in several complex variables. PZ

Complex Analysis, S(18), P. *Several Complex Variables IV: Algebraic Aspects of Complex Analysis.* Eds: S.G. Gindikin, G.M. Khenkin. Encyclop. of Math. Sci., V. 10. Springer-Verlag, 1990, 251 pp, $59.50. [ISBN: 0-387-18174-1] Four long survey articles, in expository style, on algebraic topics in the theory of functions of several complex variables: sheaves, vector bundles, deformations of complex spaces, and complex manifolds. With large bibliographies; a valuable reference work for researchers in the field. PZ

Differential Equations, T(17-18: 1), S, P. *Differential Equations: Their Solution Using Symmetries.* Hans Stephani. Ed: Malcolm MacCallum. Cambridge University Pr, 1989, xii + 260 pp, $17.95 (P); $44.50. [ISBN: 0-521-36689-5; 0-521-35531-1] A nice introduction to the theory and practice of finding and using symmetries to solve differential equations. Intuitive and practical rather than rigorous, this book is aimed at scientists and applied mathematicians. One-half of the book deals with ordinary differential equations, the other half with partial differential equations. More than 100 exercises. SP

Partial Differential Equations, T(16-17: 2), S, L. *Partial Differential Equations: Analytical Solution Techniques.* J. Kevorkian. Math. Ser. Wadsworth, 1990, xiv + 547 pp, $54.95. [ISBN: 0-534-12216-7] This book fills the vacuum between texts for a first encounter with partial differential equations and graduate texts on theory, and fills it very well. Solution techniques for linear and nonlinear equations, with discussions of physical motivations, examples, and lots of problems. Exposition is clear, problems are interesting, and the author wisely left numerical methods to be covered in other texts. JO

Partial Differential Equations, P. *Current Progress in Hyperbolic Systems: Riemann Problems and Computations.* Ed: W. Brent Lindquist. Contemp. Math., V. 100. AMS, 1989, xi + 367 pp, $41 (P). [ISBN: 0-8218-5106-3] Proceedings of the AMS-SIAM-IMS conference held at Bowdoin College in July 1988. Twenty-five papers on gas dynamics, viscoelastic materials, phase transitions, and multiphase flow in porous media. JO

Partial Differential Equations, P. *Inverse Problems in Partial Differential Equations.* Eds: David Colton, Richard Ewing, William Rundell. SIAM, 1990, 214 pp, $31.50 (P). [ISBN: 0-89871-252-1] Contains thirteen of the invited lectures at the conference on inverse scattering problems held at Humboldt State University during the summer of 1989. Concentrates on the themes of coefficient identification for parabolic equations, impedance tomography, and inverse scattering. SP

Partial Differential Equations, P. *Lecture Notes in Mathematics-1394: Nonlinear Semigroups, Partial Differential Equations and Attractors.* Eds: T.L. Gill, W.W. Zachary. Springer-Verlag, 1989, ix + 233 pp, $20 (P). [ISBN: 0-387-51594-1] Proceedings of the second symposium on this subject held at Howard University in Washington, DC in August 1987. GN

Numerical Analysis, T(18: 2). *Finite Difference Schemes and Partial Differential Equations.* John C. Strikwerda. Math. Ser. Wadsworth, 1989, xii + 386 pp, $47.95. [ISBN: 0-534-09984-X] Intended for use as an introductory graduate text for students in applied mathematics or engineering. The necessary theory of partial differential equations is developed in the text, including the basic theory of well-posed initial value problems. Includes exercises at the end of each section. GN

Functional Analysis, T(18: 1), S. *Topics in Functional Analysis and Applications.* S. Kesavan. Wiley, 1989, xii + 267 pp, (P). [ISBN: 81-224-0062-0] Paperback reprint of a 1989 hardcover (TR, October 1989). Treats Sobolev spaces, elliptic boundary value problems, semigroups of operators, and some nonlinear analysis. Four appendices, references, index. JS

Analysis, T(16-17: 1). *Lectures on Fourier Series.* L. Solymar. Oxford U Pr, 1988, x + 120 pp, $19.95 (P); $45. [ISBN: 0-19-856199-7; 0-19-856198-9] A very readable introductory text. Begins with a discussion of approximating functions, then continues with actually finding Fourier series and demonstrating their properties. Text concludes with applications to ordinary and partial differential equations. Includes many examples throughout, as well as exercises at the end of each chapter. GN

Analysis, P. *Analysis, et cetera.* Eds: Paul H. Rabinowitz, Eduard Zehnder. Academic Pr, 1990, xii + 694 pp, $69.96. [ISBN: 0-12-574249-5] The title is appropriate for a collection of papers selected for a volume that commemorates Jürgen Moser's 60th birthday—the idea being to reflect his interests in differential equations, dynamical systems, and differential geometry. AWR

Geometry, P. *The Volume of Convex Bodies and Banach Space Geometry.* Gilles Pisier. Tracts. in Math., V. 94. Cambridge University Pr, 1989, xv + 250 pp, $49.50. [ISBN: 0-521-364655] These notes present recent results (many obtained during 1984-86) that relate the volume of convex bodies in R^n and the geometry of the corresponding finite-dimensional normed spaces. They draw upon ideas from convex sets, probability, entropy numbers, and the local theory (K-convexity, cotype, and type) of Banach spaces. Contains an extensive bibliography. AWR

Geometry, S*, P, L**.** *Beyond the Third Dimension: Geometry, Computer Graphics, and Higher Dimensions.* Thomas F. Banchoff. Scientific American Library, 1990, ix + 210 pp. [ISBN: 0-7167-5025-2] A portrait of dimensions, visually rich and intellectually enriching. Slices and contours, shadows and perspective, coordinates and configuration, all beautifully illustrated with full color plates and computer-generated drawings. Offers uncommon mathematical insights into art and architecture, science and society. A wonderful gift for mathematicians to give—to their family and friends, and even more, to their students: an open door to the hidden beauty of mathematics. LAS

Optimization, P. *Lecture Notes in Mathematics-1403: Combinatorial Optimization.* Ed: B. Simeone. Springer-Verlag, 1989, v + 312 pp, $28.60 (P). [ISBN: 0-387-51797-9] Proceedings of the Centro Internazionale Matematico Estivo (CIME) International Summer School on combinatorial optimization held at Como, Italy, August 25-September 2, 1986. LC

Optimization, P. *Lecture Notes in Mathematics-1405: Optimization.* Ed: S. Dolecki. Springer-Verlag, 1989, 220 pp, $20 (P). [ISBN: 0-387-51970-X] Conference reports related to the increasing power of high speed computers: new optimization techniques being brought into the realm of possibility, and the attendent questions of complexity. AWR

Elementary Statistics, T(1, 2). *Understanding Statistics, Fifth Edition.* Lyman Ott, William Mendenhall. PWS-Kent, 1990, xvii + 727 pp. [ISBN: 0-534-92154-X] Text for general statistics for all majors; little mathematics. Surveys graphical and numerical data description, probability (not much) and sampling distributions, inferences for means, variances, proportions, regression, analysis of variance, count data. This edition adds chapters on surveys and reporting results and other additions. (*First Edition,* TR, March 1974; *Second Edition,* TR, November 1976.) TH

Statistics, P. *Lecture Notes in Statistics-59: A Calculus for Factorial Arrangements.* Sudhir Gupta, Rahul Mukerjee. Springer-Verlag, 1989, vi + 126 pp, $16.20 (P). [ISBN: 0-387-97172-6] Survey of the developments since the 1970's of the calculus for factorial arrangements. Includes an introductory chapter, derivations of characterizations for balance and orthogonality in factorial designs, applications in construction. LC

Statistics, T*(14-16), L. *Statistics: Theory and Methods.* Donald A. Berry, Bernard W. Lindgren. Brooks/Cole, 1990, xii + 763 pp. [ISBN: 0-534-09942-4] Calculus-based text covering usual topics in course for junior mathematics majors, beginning with probability. Uses many real examples for motivating and explaining. Flow is somewhat disjointed, due to examples and many topics. Worth consider-

ing. TH

Statistics, P, L. *Sampling.* Eds: P.R. Krishnaiah, C.R. Rao. Handbook of Stat., V. 6. North-Holland (US Distr: Elsevier Science), 1988, xvi + 594 pp, $125. [ISBN: 0-444-70289-X] Twenty-four chapters by 38 authors. Nineteen chapters cover general topics (e.g., systematic sampling, ratio and regression estimators) and five cover specific applications (e.g., transect sampling to assess wild life populations). TH

Statistics, T(15), L*. *Sample Design in Business Research.* W. Edwards Deming. Classics Library, Wiley, 1990, xx + 517 pp, $29.95 (P). [ISBN: 0-471-52370-4] Reprint of a 1960 text. Many stories; mathematics later in the book. Sections are "Standards of Statistical Practice," "Replicated Sampling Designs," and "Theory." TH

Programming, T(13), C. *BASIC Programming with the IBM PC, Second Edition.* Peter Mears. Ser. in Comp. Sci. Brooks/Cole, 1990, xiii + 512 pp, $38 (P). [ISBN: 0-534-12156-X] Introductory level Basic programming text assuming no prior computer experience. Includes many clear examples and exercises, and a tutorial program. Emphasis is on programming skills rather than problem solving techniques and logic. FR

Computer Systems. *Advances in Software Science and Technology, Volume 1.* Japan Soc. for Software Sci. & Tech. Academic Pr, 1990, ix + 218 pp, $49. [ISBN: 0-12-037101-4] First issue of a new monograph series produced by the Japan Society for Software Science and Technology (JSSST). It is planned on being an annual publication whose emphasis is on software research and development. *Volume 1* contains nine research contributions in such areas as formal semantics, attribute grammars, and functional languages. Future issues will be addressing a wide range of software related topics. GMS

Computer Systems, P, L*. *The Matrix: Computer Networks and Conferencing Systems Worldwide.* John W. Quarterman. Digital Pr, 1990, xxvii + 719 pp, $49.95 (P). [ISBN: 1-55558-033-5] The bible—actually, the tower of Babel—of computer networks. Layers, protocols, administration, and standards, followed by descriptions of hundreds of different (non-commercial) networks around the world—organized by region and country, each with name, address, telephone, and network address of a contact person. Sequel to an influential 1986 survey in the *Communications of the ACM.* Soon an electronic version of the book and its continuous updates will be available. LAS

Applications, P. *Selected Papers of S. Chandrasekhar, Volume 4: Plasma Physics, Hydrodynamic and Hydromagnetic Stability, and Applications of the Tensor-Virial Theorem.* S. Chandrasekhar. University of Chicago Pr, 1989, xiii + 585 pp, $29.95 (P); $79.95. [ISBN: 0-226-10097-9; 0-226-10096-0] Thirty-five papers by author and coauthors from 1954 through 1971. In three parts: plasma physics, hydrodynamic and hydromagnetic stability, and tensor virial theorem and its applications. Volume four of six. JDEK

Applications, P. *Selected Papers of S. Chandrasekhar, Volume 3: Stochastic, Statistical, and Hydromagnetic Problems in Physics and Astronomy.* S. Chandrasekhar. University of Chicago Pr, 1989, xiv + 642 pp, $34.50 (P); $91.95. [ISBN: 0-226-10095-2; 0-226-10094-4] Forty-six papers by author and co-authors from 1941 through 1958. In four parts: dynamical friction and Brownian motion, statistical problems in astronomy, statistical theory of hydrodynamic and hydromagnetic turbulence, and hydromagnetic problems in astrophysics. Volume three of six. JDEK

Applications (Economics), P. *Recursive Methods in Economic Dynamics.* Nancy L. Stokey, Robert E. Lucas, Jr. Harvard U Pr, 1989, xviii + 588 pp, $37.50. [ISBN: 0-674-75096-9] The idea of the authors is to develop the substantial mathematical tools of recursive methods: metric space concepts for optimization, measure theory and Markov theory for probabilistic techniques, fixed point theorems for the study of stability. After presentation of the mathematical tools, applications are taken up. This is a high-powered book that comes with some high-powered endorsements that will be tough to live up to. Only time will tell if it achieves the stature of Hick's *Value and Capital,* but it does break some new ground in its full development of recursive methods as a unifying thread. AWR

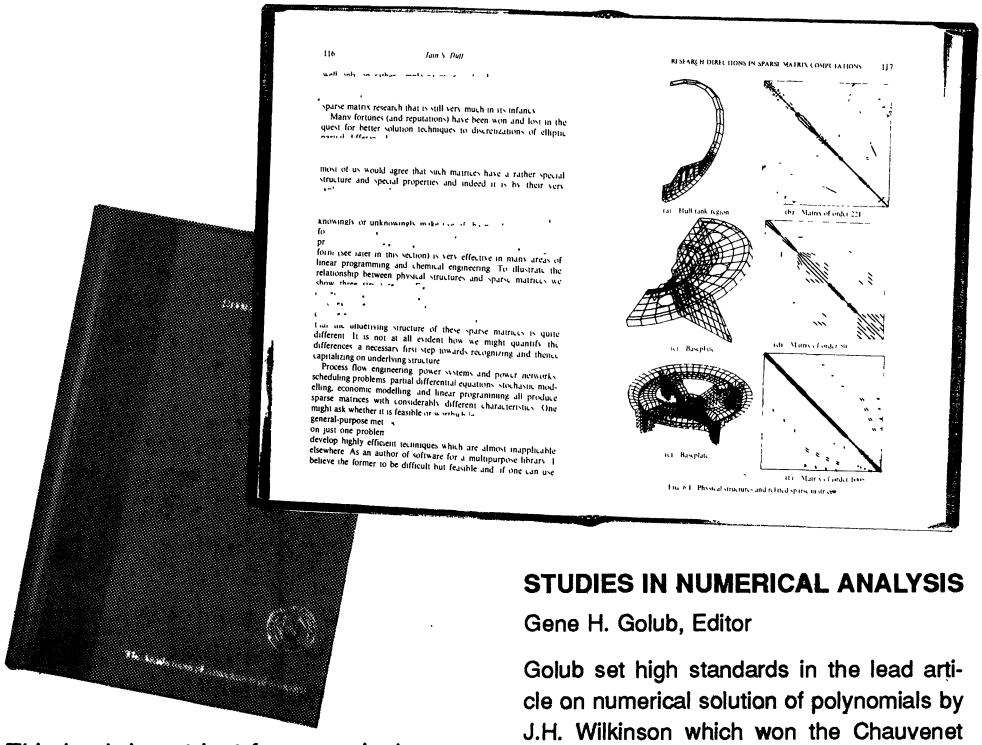
Applications (Engineering), P. *Computational Aspects of VLSI Design with an Emphasis on Semiconductor Device Simulation.* Ed: Randolph E. Bank. Lect. in Appl. Math., V. 25. AMS, 1990, viii + 190 pp, $53. [ISBN: 0-8218-1132-0] Monograph contains reprints of the twelve invited papers presented at the 18th SIAM Summer Seminar on Applied Mathematics. The papers present advanced research on the use of numerical simulation for the design of VLSI chips, allowing engineers to test and evaluate chip design before going into the more costly and time-consuming prototyping phase. This shortens the overall design cycle considerably. GMS

Applications (Engineering), P. *Lecture Notes in Engineering-52: Slope Analysis Using Boundary Elements.* Y.S. Jiang. Springer-Verlag, 1989, iv + 176 pp, $27 (P). [ISBN: 0-387-51625-5] Slope stability is a classical topic in soil and rock mechanics. These notes study slope failure due to slip failure: a mass of soil or rock sliding over a surface. The boundary element method is studied as a tool in slope failure analysis. LC

Reviewers

DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; LC: Laura Chihara, St. Olaf; TH: Timothy Hesterberg, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; GN: Gail Nelson, Carleton; JO: Jeff Ondich, St. Olaf; SP: Samuel Patterson, Carleton; FR: Flauren Ricketts, Macalester; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; LAS: Lynn Arthur Steen, St. Olaf; PZ: Paul Zorn, St. Olaf.

STUDIES IN NUMERICAL ANALYSIS



This book is not just for numerical analysts, it is accessible to anyone with some mathematical foundation who wants to taste selections from a numerical analysis menu. The papers do a good job in motivating and introducing their subjects, and some provide a nice historical background. I recommend the articles in this book for anyone interested in learning more about the current research interests in numerical analysis. They may excite their readers to investigate further this relevant field of mathematics.

David C. Arney in *Mathematics and Computer Education*

STUDIES IN NUMERICAL ANALYSIS

Gene H. Golub, Editor

Golub set high standards in the lead article on numerical solution of polynomials by J.H. Wilkinson which won the Chauvenet Prize as the best piece of mathematical exposition to appear in 1985. These papers describe the wide range of research in numerical analysis. A variety of problems and of computational tools is treated. Some of the computations require only a hand-held calculator, while others require a powerful computer.

422 pp., 1985, ISBN 0-88385-126-1

List: $46.50 MAA Member: $34.50

Catalog Number MAS-24

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

Cambridge University Press

Volterra Integral and Functional Equations
G. Gripenberg, S.O. Londen, and O. Staffans

Convincingly demonstrates that the theory of Volterra equations exhibits a rich variety of features not present in the theory of ordinary differential equations.

Encyclopedia of Mathematics and its Applications 34
1990 701 pp. 37289-5 Hardcover $99.50

The Mathematics of Projectiles in Sport
Neville de Mestre

A comprehensive and rigorous account of the techniques applied to the motion of projectiles. The final section of the book is an enlightening collection of sporting applications, ranging from the high jump to frisbees and soccer to table tennis.

Australian Mathematical Society Lecture Series 6
1990 174 pp. 39857-6 Paper $22.95

Introduction to Lattices and Order
B.A. Davey and H.A. Priestley

Addresses the importance of ordered structures such as sets and lattices. Recognizes the increased importance of ordered structures in theoretical computer science and all areas of discrete mathematics.

1990 150 pp. 36584-8 Hardcover $49.50
36766-2 Paper $18.95

Boojums All the Way Through
Communicating Science in a Prosaic Age
N. David Mermin

This collection of nontechnical writings in and about phyics deals in a variety of ways with the process of communicating modern physics to both physicists and nonphysicists. With wry humor and a style for liveliness, Mermin demonstrates his ability to convey complex ideas with startling simplicity.

1990 308 pp. 38231-9 Hardcover $49.50
38880-5 Paper $16.95

Basic Hypergeometric Series
George Gasper and Mizan Rahman

An authoritative account of basic hypergeometric series and their applications. Provides explicit and detailed information about summation, transformation and expansion formulas, and contour integrals.

Encyclopedia of Mathematics and its Applications 34
1989 350 pp. 35049-2 Hardcover $59.50

Mathematics and Cognition
A Research Synthesis by the International Group for the Psychology of Mathematics Education
Pearla Nesher and Jeremy Kilpatrick

A distinguished collection describing research findings that shed light on the learning of mathematics from early arithmetic to high levels of algebra and geometry.

International Commission on Mathematical Instruction Study Series
1990 180 pp. 36608-9 Hardcover $39.50
36787-5 Paper $14.95

Continuous and Discrete Modules
Saad M. Mohamed and Bruno J. Muller

A complete and up-to-date account of the subject of continuous and discrete modules that at the same time gives a unified picture of the theory.

London Mathematical Society Lecture Note Series 147
1990 126 pp. 39975-0 Paper $22.95

Lectures on Minimal Surfaces
Volume 1
Johannes C.C. Nitsche

A broad survey of parametric minimal surface in Euclidean space that extends from the classical beginnings to the current situation.

1989 563 pp. 24427-7 Hardcover $135.00

Representations of Nilpotent Lie Groups and Their Applications
Part I
Frederick P. Greenleaf and Laurence Corwin

A survey of harmonic theory developed since the pioneering work of Kirillov in 1958. Provides not only a modern account of all topics necessary for current research, but also includes many computed examples.

Cambridge Studies in Advanced Mathematics
1990 269 pp. 36034-X Hardcover $69.50

Now in paperback...
Winner of the Outstanding Academic Books/Nonprint Materials 1988-89 Award
— *Choice Magazine*

Fourier Analysis
T.W. Korner

"...an extraordinary and very attractive book... I would like to see the book on the desk of every pure mathematician with an interest in analysis, and of every teacher of applied mathematics whose work involves analysis." — SIAM Review

1988 591 pp. 38991-7 Paper $34.50

At bookstores or from
CAMBRIDGE UNIVERSITY PRESS
40 West 20th Street, New York, NY 10011. Call toll-free 800-872-7423 outside NY State. 800-227-0247, NY State only.
MasterCard/VISA accepted. Prices subject to change.

GEOMETRY & TOPOLOGY

FIRST CONCEPTS OF TOPOLOGY: The Geometry of Mappings of Segments, Curves, Circles, and Disks

W.G. Chinn and N.E. Steenrod

This clear and winning little book, for readers willing to come to genuine grips with the idea of a mathematical proof, presents topology . . . as mathematicians see it. The parlor tricks are gracefully alluded to here and there, but they are distinctly for after hours. The center of interest is the stuff itself: powerful notions of set theory are . . . exploited to define open sets and their coverings, and from them [to prove] the key theorems. .. One cannot any longer doubt that a single stroke of a knife exists that divides any irregular ham sandwich so that ham and both bread slices can be shared with perfect fairness by two consumers.

Philip Morrison in *Scientific American*

Contains over one hundred and fifty problems and solutions.

160 pp., 1966, ISBN 0-88385-618-2

List: $10.00 MAA Member: $8.00

Catalog Number NML-18

FROM PYTHAGORAS TO EINSTEIN

K.O. Friedrichs

Starting with area and dissection proofs of the Pythagorean theorem, Friedrichs proceeds gently, step by step, to cover vectors, coordinates, elastic and inelastic impacts, and relativistic space-time–ending with a derivation of a contemporary formula rivaling the Pythagorean theorem in fame, $E = mc^2$.

88 pp., 1965, ISBN 0-88385-616-6

List: $9.50 MAA Member: $7.60

Catalog Number NML-16

GEOMETRY REVISITED

H.S.M. Coxeter and S.L. Greitzer

After seeing a proof that a segment has a unique midpoint, students may wonder whether there are any interesting theorems in geometry. If either they or their teachers get a hold of this book, they will learn that there are many beautiful and nontrivial theorems in geometry. Among those found here are the theorems of Ceva, Menelaus, Pappus, Desargues, Pascal, and Brianchon. A nice proof is given of Morley's remarkable theorem on angle trisectors. The transformational point of view is emphasized: reflections, rotations, translations, similarities, inversions, and affine and projective transformations. Many fascinating properties of circles, triangles, quadrilaterals, and conics are developed.

193 pp., 1967, ISBN 0-88385-619-0

List: $11.75 MAA Member: $9.50

Catalog Number NML-19

The very lucid presentation takes the reader from the elementary problems of plane Euclidean geometry to the fundamental concepts of non-Euclidean geometry . . . The book is rich in remarkable facts and thereby is very effective in promoting the significance and the value of geometry in mathematical teaching, a promotion which is very necessary . . . The always original developments use very simple tools . . . and soon proceed to higher configurations.

K. Strubecker in *Mathematical Reviews*

Order from:



The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036
(202) 387-5200

When was the last time a computer program helped you think about mathematics?

Call or write for our latest software catalog.
For PC or Macintosh.

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

Several Complex Variables

Proceedings of the Mittag-Leffler Institute, 1987–1988

Edited by John E. Fornaess

There was a special year devoted to the topic of several complex variables at the Mittag-Leffler Institute in Stockholm, Sweden, and this volume contains the resulting survey papers and research papers. The work covers a broad spectrum of recent developments in the field.

The contributors to the volume are H. Alexander; F. Almgren; E. Almar; M. Andersson; E. Bedford; J. Belanger; S. Bell; B. Berndtsson; U. Cegrell; C.H. Chang and H.P. Lee; J.P. Demailley; P. Dolbeault; A. Dor; F. Forstneric; B. Gaveau, M. Okada, and T. Okada; R. Greene and S. Krantz; A. Iordan; C. Laurent-Thiebaut and J. Leiterer; L. Lempert; I. Lieb and M. Range' L. Qi-King; P. Manne; A. Noell; M. Passare; J. Riihentaus; J.P. Rosay and W. Rudin; R. Saerens and W. Zame; A. Sergeev; N. Sibony; E.L. Stout; F. Treves; S. Webster; H.H. Wu; and A. Zeriahi.

Mathematical Notes, 38

Paper: $32.50 ISBN 0-691-08579-X

In Japan, order from United Publishers Services



AT YOUR BOOKSTORE OR

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900

ORDERS 800-PRS-ISBN (777-4726)

INTERNATIONAL MATHEMATICAL OLYMPIADS

The International Mathematical Olympiads, 1959-1977 and 1978-1985 provide a compilation of problems and solutions of arresting ingenuity; all accessible to secondary school students. The alternative solutions are particularly interesting because they show that there are many ways to solve a problem.

INTERNATIONAL MATHEMATICAL OLYMPIADS, 1959-1977

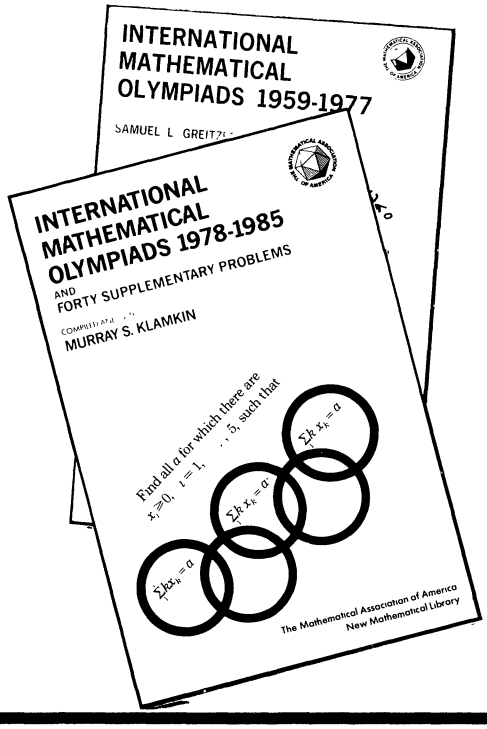
Compilation and solutions by S.L. Greitzer

116 problems

204 pp., 1978, ISBN 0-88385-627-1

List: $11.50 MAA Member: $9.50

Catalog Number NML-27



INTERNATIONAL MATHEMATICAL OLYMPIADS , AND FORTY SUPPLEMENTARY PROBLEMS, 1978-1985

Compilation, solutions, and 40 additional problems by Murray S. Klamkin

88 problems in all

150 pp., 1986, ISBN 0-88385-631-X

List: $12.95 MAA Member: $10.50

Catalog Number NML-31

This sequel of problems used in the International Mathematical Olympiad certainly matches the description 'meant to challenge.' The problems contained in this publication are the most challenging I have seen. For the serious problem solver, this is an excellent collection of truly challenging problems designed to give instruction in the use of sophisticated methods and modes of attack.

Richard J. Wylie in *The Mathematics Teacher*

Both International Olympiad books available as a package:

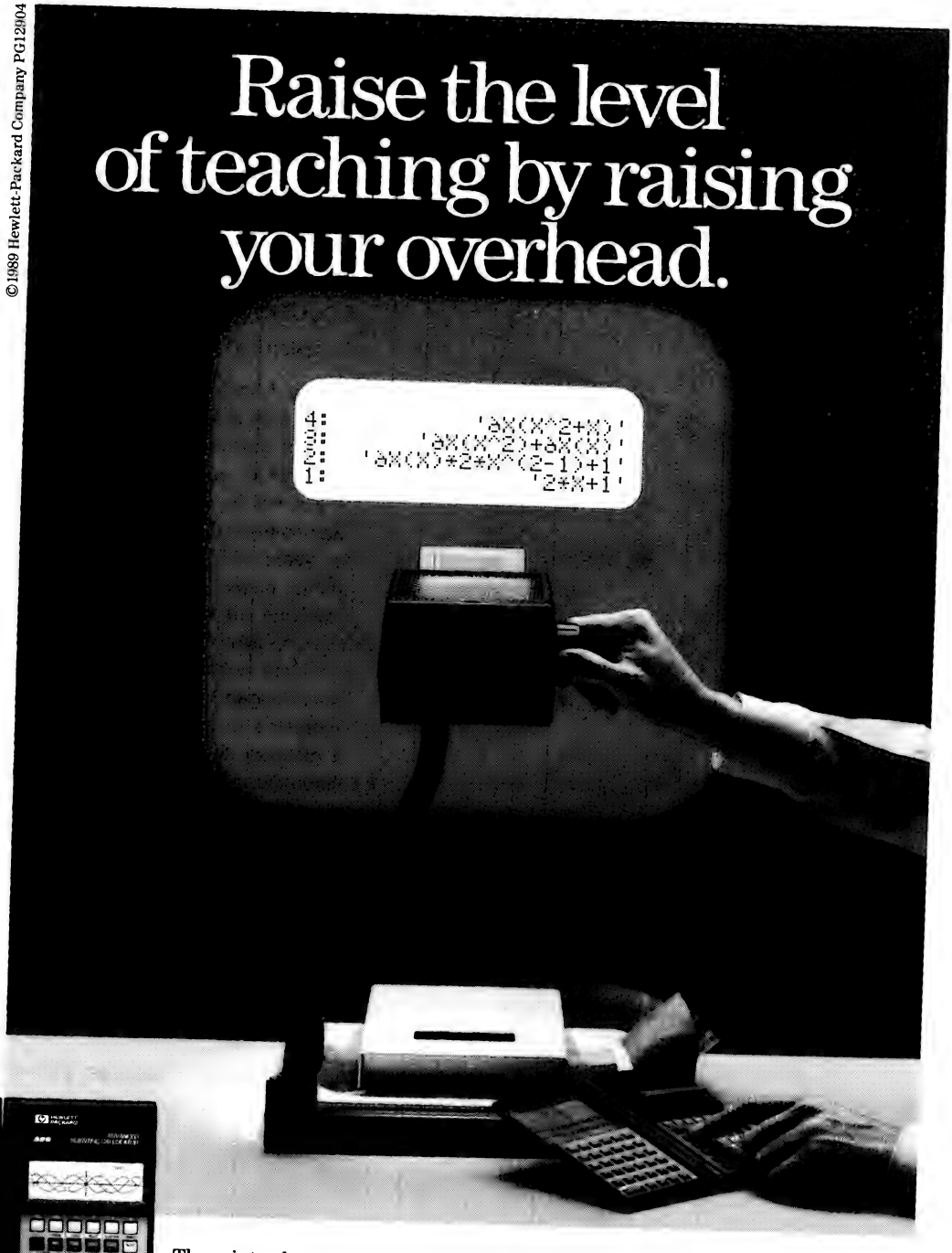
List: $21.00 MAA Member: $17.50

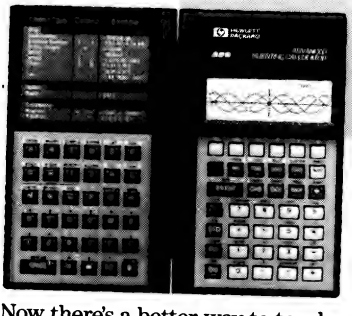
Catalog Number IMO

Order from:
The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

© 1989 Hewlett-Packard Company PG12904

Raise the level of teaching by raising your overhead.





Now there's a better way to teach algebra and calculus in the classroom. In fact, two ways.

First, introduce your students to the HP-28S. It's the only calculator that offers symbolic algebra and calculus.

Then, introduce yourself to the overhead display for the HP-28S. It allows you to project your calculations on an HP-28S for everyone in the classroom to see.

A scholastic offer for you.

If your department or students purchase a total of 30 HP-28S calculators, we'll give you a classroom overhead display for the HP-28S absolutely free. (A $500 retail value.) Plus, your very own HP-28S calculator free. (A $235 retail value.)

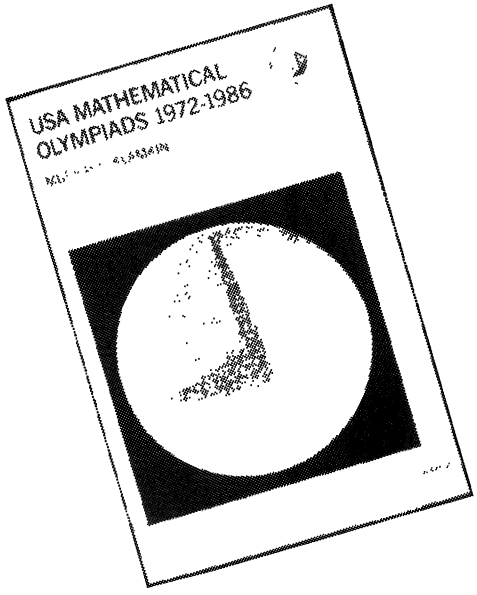
To learn more, and get free curriculum materials, call (503) 757-2004 between 8am and 3pm, PT. Offer ends October 31, 1989.

There is a better way.

HEWLETT PACKARD

Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330

USA MATHEMATICAL OLYMPIADS



Every year 100 of the most mathematically talented high school students in the country compete in the USA Mathematical Olympiad (USAMO). The USAMO is the third stage of a three-tiered mathematical competition for high school students in the United States and Canada that begins with the AHSME taken by over 400,000 students, continues with the American Invitational Mathematics Exam involving 2,000 students, and culminates in the 100-contestant USAMO.

USA MATHEMATICAL OLYMPIADS 1972-1986, PROBLEMS AND SOLUTIONS

Compiled by Murray S. Klamkin

People delight in working on problems "because they are there, " for the sheer pleasure of meeting a challenge. This is a book full of such delights. In it, Murray S. Klamkin brings together 75 original USA Mathematical Olympiad (USAMO) problems for years 1972-1986, with many improvements, extensions, finger exercises, open problems, references and solutions, often showing alternative approaches. The problems are coded by subject and solutions are arranged by subject as an aid to those interested in a particular field. Contains a glossary of frequently used terms and theorems, and a comprehensive bibliography with items numbered and referred to in brackets in the text. The problems are intriguing and the solutions elegant and informative. Students and teachers will enjoy working these challenging problems. Indeed all those those who are mathematically inclined will find many delights and pleasant challenges in this book.

180 pp., 1988, ISBN-0-88385-634-4

List: $13.50 MAA Member: $12.50

Catalog Number NML-33

Order from:

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200



PLANT THE SEEDS TO HIGHER LEARNING . . .

Being first is almost second nature at McGraw-Hill. Whether your students are enrolled in developmental math, precalculus, or calculus, McGraw-Hill's texts will inspire your students to learn.

NEW FOR 1991

The Streeter Series

BASIC MATHEMATICAL SKILLS WITH GEOMETRY
Streeter/Hutchison/Hoelzle

BEGINNING ALGEBRA, Form B
Streeter/Hutchison/Hoelzle

INTERMEDIATE ALGEBRA, Form B
Streeter/Hutchison/Hoelzle

COLLEGE ALGEBRA WITH TRIGONOMETRY
Rees/Sparks/Rees

PRECALCULUS: A Bridge to Calculus
Runyan/Runyan

EXPLORATIONS IN CALCULUS WITH A COMPUTER ALGEBRA SYSTEM
Small/Hosack

DISCOVERING CALCULUS WITH THE HP-28s AND THE HP-48sx
Smith/Minton

DISCRETE MATH AND ITS APPLICATIONS, 2/e
Rosen

DIFFERENTIAL EQUATIONS WITH APPLICATIONS, 2/e
Simmons

INTRODUCTION TO PARTIAL DIFFERENTIAL EQUATIONS, 2/e
Pinsky

FUNCTIONAL ANALYSIS, 2/e
Rudin

MATRIX THEORY
Goldberg

BRIDGE TO ABSTRACT MATHEMATICS, 2/e
Morash



For more information, please contact your local McGraw-Hill representative or write: McGraw-Hill College Division, Comp Processing and Control, P.O. Box 448, Hightstown, NJ 08520.

New. . .

Writing Mathematics Well, by Leonard Gillman

64 pp., 1987, ISBN-0-88385-443-0
Catalog Number - WMW List: $6.00 MAA Member: $4.50

Good writing conveys more than the author originally had in mind, while poor writing conveys less. Well-written papers are more quickly accepted and put into print and more widely read and appreciated than poorly written ones—and for notes, monographs, and books the quality of writing is more important than it is for papers.

In **Writing Mathematics Well,** Leonard Gillman tells his readers how to develop a clear and effective style. All aspects of mathematical writing are covered, from general organization and choice of title, to the presentation of results, to fine points on using words and symbols, to revision, and finally, to the mechanics of putting your manuscript into print. No book can by itself make you a better writer, but this one will alert you to the opportunities for better and more forceful writing. It does this both by precept and by example.

A book to be read for its sharpness and wit as well as for enlightenment, **Writing Mathematics Well** should be on the shelf of anyone who writes or intends to write mathematics. It will amuse and delight the already careful writer and it will help reform and refine the sensibilities of those who may be somewhat careless about their writing.



Order from
The Mathematical Association of America
1529 Eighteenth St. NW
Washington, D.C. 20036



Carus Mathematical Monograph No. 20

THE GENERALIZED RIEMANN INTEGRAL

Robert M. McLeod
275 + xiii pages. Hardbound.
List: $24.95 MAA Member: $22.00

Until recently the most powerful and beautiful tools of integration theory have been accessible only to the privileged few whose studies extended through the Lebesgue integral. Now a new integral, a generalization of the familiar Riemann integral, has been discovered which has all the power and range of the Lebesgue integral but which can be readily understood by anyone who has studied calculus through the sophomore level.

The Generalized Riemann Integral is the first book-length presentation of this exciting new development in integration theory. Because of the clarity and organization of the exposition and the inclusion of exercises designed to actively engage the reader in the material, it is eminently suitable for use as a textbook at the advanced undergraduate or beginning graduate level. Furthermore, because it presents within a single volume results which were previously scattered throughout many research publications, it will undoubtedly also be of considerable interest to specialists in integration theory.

List of Contents: Definition of the Generalized Riemann Integral • Basic Properties of the Integral • Absolute Integrability and Convergence Theorems • Integration of Subsets of Intervals • Measurable Functions • Multiple and Iterated Integrals • Integrals of Stieltjes Type • Comparison of Integrals • Appendix: Solutions of In-Text Exercises

THE MATHEMATICAL ASSOCIATION OF AMERICA

Order From:
Mathematical Association of America
1529 Eighteenth Street NW
Washington DC 20036

Innovation and Quality — HBJ Mathematics for 1991

When you asked for a sensible, no-frills, mainstream calculus textbook at a reduced cost for your students, we heard you.

CALCULUS
One and Several Variables
ROBERT ELLIS and DENNY GULICK
both of *University of Maryland*
Hardcover, 1991, Just Published!

SINGLE-VARIABLE CALCULUS WITH DISCRETE MATHEMATICS
JOHN A. FEROE and
CHARLES STEINHORN
both of *Vassar College*
Hardcover, 1991

ELEMENTARY LINEAR ALGEBRA WITH APPLICATIONS
Second Edition
RICHARD O. HILL, JR.
Michigan State University
Hardcover, 1991
Available in November for spring classes!

ABSTRACT ALGEBRA
DENNIS KLETZING, *Stetson University*
Hardcover, 1991

FINITE MATHEMATICS
WILLIAM OWEN and FRED CUTLIP
both of *Central Washington University*
Hardcover, 1991

CALCULATOR ENHANCEMENT FOR SINGLE-VARIABLE CALCULUS
JOHN KENNELLY and
JAMES H. NICHOLSON
both of *Clemson University*
1991, Just Published!

CALCULATOR ENHANCEMENT FOR MULTIVARIABLE CALCULUS
JAMES A. RENEKE, *Clemson University*
1991, Just Published!

CALCULATOR ENHANCEMENT FOR DIFFERENTIAL EQUATIONS
THOMAS G. PROCTOR
Clemson University
1991, Just Published!

CALCULATOR ENHANCEMENT FOR LINEAR ALGEBRA
DON LA TORRE, *Clemson University*
1991, Just Published!

CALCULATOR ENHANCEMENT FOR INTRODUCTORY STATISTICS
IRIS FETTA, *Clemson University*
1991, Just Published!

HBJ

HARCOURT BRACE JOVANOVICH, INC.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648
(708) 647-8822

Springer for Mathematics

Celebrating the 75th Anniversary of the MAA

G.A. Edgar, The Ohio State University, Columbus, OH

Measure, Topology, and Fractal Geometry

Intended to provide the mathematics necessary for the study of fractal geometry. It includes background material on metric topology and measure theory. It covers topological dimension and fractal dimension (in particular Hausdorff dimension). It contains a complete discussion of self-similarity, and the more general "graph self-similarity."
1990/248 pp., 96 illus./Hardcover/$29.95
ISBN 0-387-97272-2
Undergraduate Texts in Mathematics

J. Stillwell, Monash University, Victoria, Australia

Mathematics and Its History

This concise, unified view of mathematics and its historical development identifies reasons for the emergence of the main fields of modern mathematics, and explains connections between them by tracing the course of a few mathematical themes from ancient times through the 20th century.
Students can expect to add to their mathematical knowledge as well as gain a new perspective on what they already know, and those wishing to explore the classics for themselves will benefit from Stillwell's use of modern terms to express his interpretation of classical mathematics.
1989/371 pp., 163 illus./Hardcover/$49.80
ISBN 0-387-96981-0
Undergraduate Texts in Mathematics

S. Lang, Yale University, New Haven, CT

Undergraduate Algebra

Second Edition

Written in the author's clear style, this new second edition includes some additional topics, such as Sylow groups, Mason's theorem on polynomials and the analogous abc conjecture over the integers, and expands on other topics, including symmetric polynomials and field theory. Many new exercises and concrete examples of major unsolved problems in algebra, written in a language accessible to undergraduates, have been added to illuminate the vitality of mathematics.
1990/app. 368 pp., 3 illus./Hardcover/$39.80
ISBN 0-387-97279-X
Undergraduate Texts in Mathematics

K.L. Chung, Stanford University, Stanford, CA

Elementary Probability Theory with Stochastic Processes

Third Edition, 3rd Printing, 1990

Here is an introductory text to probability and some of its applications. The content is adequate for a self-contained elementary course for undergraduates or as a prerequisite to more advanced courses on the subject and allied fields such as statistics, operations research, and engineering.
1979/325 pp., 36 illus./Hardcover/$39.95
ISBN 0-387-90362-3
Undergraduate Texts in Mathematics

M.A. Armstrong, University of Durham, England

Basic Topology

3rd Printing, 1990

From the reviews:
"The book is very good, its material sensibly chosen for an undergraduate course, and with a clear exposition. It has good and plentiful illustrations - for the author, topology is above all a geometrical subject - and many exercises."
— Mathematical Gazette
1983/250 pp./Hardcover/$34.00
ISBN 0-387-90839-0
Undergraduate Texts in Mathematics

To Order:

Call Toll-Free: **1-800-SPRINGER** (in NJ call 201-348-4033), or send payment plus $2.50 for postage and handling to: **Springer-Verlag New York, Inc.,** Attn: S. Klamkin - Dept. S563, 175 Fifth Avenue, New York, NY 10010. *NY, NJ and CA residents please include state sales tax.*

Call or write for information on textbook examination copies!



Springer-Verlag

New York • Berlin • Heidelberg • Vienna • London • Paris • Tokyo • Hong Kong • Barcelona

THE MATHEMATICAL ASSOCIATION OF AMERICA

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 8 October 1990



GEOMETRY

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Macalester College, St. Paul, MN 55105. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

Special Geometry Issue

The *American Mathematical Monthly* is pleased to present, in this issue, a tribute to geometry, one of the oldest-newest parts of our subject, in the form of six articles by some of its leading practitioners.

The poem below, which was part of Professor Chern's article "What is geometry?," is displayed here because it beautifully sets the tone of the issue.

—Herbert S. Wilf, for the editors

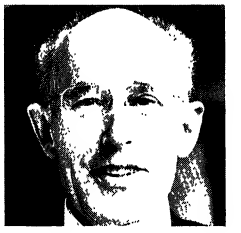
A Poem

Physics and geometry are one family
Together and holding hands they roam to the
limits of outer space.
Black hole and monopole exhaust the secret of myths;
Fiber and connection weave to interlace
the roseate clouds.
Evolution equations describe solitons;
Dual curvature defines instantons.
Surprisingly, Math. has earned its rightful place
for man and in sky;
Fondling flowers with a smile—just wish
nothing is said!

(translated by Harry Wang, from the original on p. 686.)

Convexity

MARCEL BERGER, IHES, *Bures-sur-Yvette, France*



MARCEL BERGER: Born Paris, France, 1927. After Ph.D. on "Holonomy groups of Riemannian manifolds" under A. Lichnerowicz, was professor at the universities of Strasbourg, Nice and Paris 1955–1974. From 1974–1985 Director of Research with the CNRS (Centre National de la Recherche Scientifique). Since 1985, Director of the IHES (Institut des Hautes Etudes Scientifiques). Visited a full year MIT 1956–57 and University of California, Berkeley 1961–62. President of the French Mathematical Society 1979–81. Corresponding member of the French Academy of Sciences since 1982. Rademacher lecturer (Univ. of Pennsylvania) 1981. Editor and managing editor of various mathematical journals, managing editor of the yellow Springer *Grundlehren des Mathematischen Wissenschaften*. Published around 45 papers on Riemannian geometry and three books (all with Springer): *Geometry I and II*, (with Gostiaux) *Differential Geometry: Manifolds, Curves and Surfaces*, and (with Berry, Pansu and St. Raymond) *Problems in Geometry*.

Convexity, as we shall see, is a very old topic which can be traced at very least to Archimedes. It has more or less always been in favor, and now it is making a very strong comeback. This can be attributed in part to the rise of linear programming and the computer era starting from the '60s. But the geometric method in analysis has come up with wonderful results including some spectacular ones on convex bodies.

At the same time, convexity is an extremely simple and natural notion. So we think the reader will readily appreciate what follows. Interesting in itself, it will also illustrate some facts about mathematics, facts that are more or less classical, but always important to realize, paradoxical though they may be. First, questions or problems arise that are very simple to formulate (as in number theory) but to which the answers are either still unknown or have been found only very recently, often using very hard techniques from other parts of mathematics. The second fact is that for elementary geometric problems formulated in our ordinary 2 or 3-dimensional space, one is forced to use abstraction and, among other things, to "go to the infinite" or to higher dimensional spaces. Finally, intuition is sometimes quite misleading.

Because of the restricted length of this article, I cannot be exhaustive. Indeed I had to select only a few topics. Selection was based on naturalness and simplicity, my own taste, and illustration of the facts I have just mentioned. Standard topics which really could not be included receive a passing mention in the last section. The material has been organized as follows.

1. Convexity is a natural notion; historical examples
2. Rigorous definitions; examples
3. The John-Loewner ellipsoid; applications
4. Convex functions; examples and applications
5. Polytopes: four "elementary" problems
6. Two algebraic operations on the set of all convex bodies: duality and addition

7. Topology in the set of all convex bodies: intuition is dangerous
8. A brief look at some topics in convexity

To prepare the present text I received help from, among others, Peter Gruber and Vitali Milman. It is a pleasure for me to thank them here.

1. Convexity is a natural notion; historical examples

In art, the words *concave* and *convex* are commonly used, as shown in the following comment on a modern sculpture:

> This sculpture reflects the influence of Cubism in its optical juxtaposition of interchangeable *concave and convex forms*, and in its use of a void to express a mass (from *Treasures of the Israel Museum, Jerusalem*, Geneva, 1985).



Woman Combing Hair (Bronze)—Alexander Archipenko (1887–1964)
(Courtesy the Israel Museum, Jerusalem)

The same applies to anatomy textbooks:

> Menisci or half-moon fibro-cartilages. *Arranged in this manner, the glenoid cavities do not adapt to the femur condyles. They fit together by means of menisci or half-moon fibro-cartilages placed between the tibia and the femur.*

Fibro-cartilages, like glenoid cavities, are external or internal. Each is a triangular prismatic blade curved into a crescent.

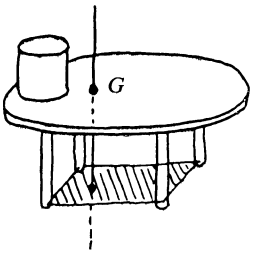
They are seen to possess: a concave *upper face, connected with the femur condyles; an underside, in contact with the periphery of the corresponding glenoid cavity; an outer or peripheral face (the base of the prism), which is* convex, *very thick, adhering to the articular capsule; an inner or central edge, which is* concave, *trenchant, and whose* concavity *faces the centre of the glenoid cavity.* ... (translation from a classic French book on anatomy by Rouvière).

Note that convexity and concavity appear to be such commonsense notions that neither is defined in texts on art or anatomy!

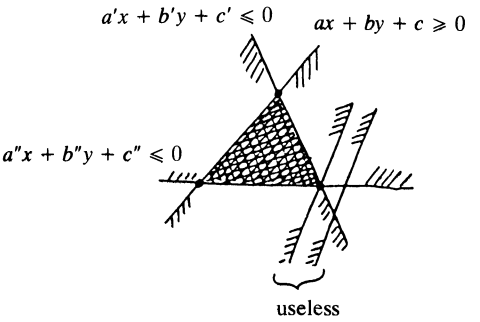
Archimedes (circa 250 B.C.) explicitly stated that the inner curve of the figure below is shorter than the outer one, if the inner one is convex. This is obviously false if it is not.



Poinsot (circa 1800) looked into convexity when he studied statics, stating, for example, that to ensure stability for the table below one needs the vertical line through the center of gravity to intersect the supporting plane inside the convex envelope (see section 2) of the set made up by the legs of the table.



Fourier also studied statics around the same time and was led to study simultaneous linear inequalities like these:



He was smart enough to realize the need to determine which ones were really relevant (see section 5). This was the origin of linear programming which began to

develop in the '60s. It consists for example in asking:

Find the maximum of $y = x_1 - x_2 + 2x_3$ if

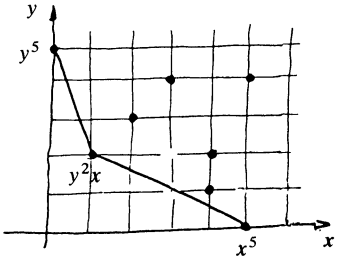
$$\begin{cases} x_1 + x_2 + 3x_3 + x_4 \leq 5 \\ x_1 + x_3 - 4x_4 \leq 2 \\ x_1 \geq 0. \end{cases}$$

A large variety of computer programs are available to solve such problems. The remark announced above is that most of the interest in linear programming (through computers) lies in working with very large numbers of inequalities and variables. Hence the need (by no means a mathematician's luxury) to work and develop intuition in high-dimensional spaces.

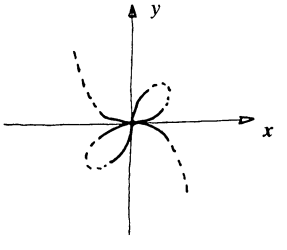
Back in the 1720s Newton was already using convexity in a basic manner to solve the problem of finding the local shape of a real algebraic plane curve at a singular point, however complicated. His solution was quite complete, as the following example shows:

$$2y^5 + y^4x^3 - 7y^4x^5 + 3y^3x^2 - y^2x^4 - 5y^2x + yx^4 + x^5 = 0.$$

The singularity to be studied is at the origin (0, 0). If this is not the case adjust it by suitable translation of coordinates. Now put a dot at (m, n) in the lattice $\mathbf{N} \times \mathbf{N}$ of integral points in the plane $\mathbf{R}^2$ for every $a_{m,n}x^m y^n$ in the equation defining the curve, for which $a_{m,n} \neq 0$:



Then draw the convex envelope (paying attention only to the west and south parts) of this set of dots, thereby obtaining a number of segments. For each segment, extract the two terms that correspond to the end points of that segment from the original equation of the curve. Draw each of these two-term curves locally. Newton's theorem is that, whatever the other dots, the union of the preceding curves drawn gives the *local* shape of the total curve. See [**14**], [**36**] for more details and the general theory.



three branches

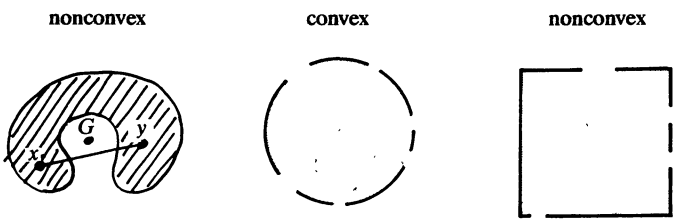
$$2y^5 - 5y^2x = 0 \qquad x^5 - 5xy^2 = 0$$
$$\Downarrow \qquad\qquad \Downarrow$$
$$2y^3 - 5x = 0 \qquad x^4 - 5y^2 = 0$$
$$\Downarrow \qquad\qquad \Downarrow$$
$$5x = 2y^3 \qquad x^2 = \sqrt{5}\,y \text{ and } x^2 = -\sqrt{5}\,y$$

$$(x^5 - 5xy^2 + 2y^5 = 0)$$

2. Rigorous definitions; examples

In chapters 2 to 6 we will work exclusively in the linear space $\mathbf{R}^d = \{x = (x_1, \ldots, x_d)\colon x_i \in \mathbf{R}\}$. According to the nature of the problem that we will be studying, we consider it as *endowed* or *not* with its canonical Euclidean metric $d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_d - y_d)^2}$.

A subset K of $\mathbf{R}^d$ is said to be *convex* if for every x and y in K the segment $[x, y]$ of $\mathbf{R}^d$ whose ends are x and y is all contained in K: $[x, y] \subset K$. The pictures below show convex and nonconvex sets. Because the open disk $x^2 + y^2 < 1$, plus *any* part of its boundary (the circle $x^2 + y^2 = 1$) is convex, we will always work with *open* or *closed* convex sets. Note that things are different for the square. Recall, if need be, that a subset K of $\mathbf{R}^d$ is said to be *open* if any of its points x is the center of an open ball $B(x, r) = \{y\colon d(x, y) < r\}$ $(r > 0)$ entirely contained in K. It is said to be *closed* if its *complement* $\mathbf{R}^d \setminus K$ is open. It is equivalent to say that the limit x of every convergent sequence $\{x_i\}$ is still in K.

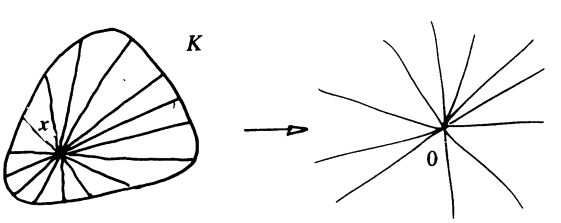


Three immediate properties of convex sets are the following:

The first was already known to Archimedes: if K is a plate in $\mathbf{R}^2$ or a body in $\mathbf{R}^3$, of some material, then it contains its center of mass (center of gravity). Note that constant density is not required.

The second property: look at the two following metrics on a subset K of $\mathbf{R}^d$. The first, $d_{\mathbf{R}}$ is the so-called *induced* one (from the Euclidean structure of $\mathbf{R}^d$) $d_{\mathbf{R}^d}(x, y) = d(x, y)$ for any $x, y \in K$ and the Euclidean metric d on $\mathbf{R}^d$. The second, called *intrinsic*, *denoted* by d_K, is defined as the infimum $d_K(x, y)$ of the length in $\mathbf{R}^d$ of all curves from x to y which lie entirely within K. Then it is easy to see that K is convex if and only if d_K is identical with $d_{\mathbf{R}}$.

The third property belongs to the realm of algebraic topology, which studies the properties of objects that depend only on maps preserving the topology, in particular those properties that are invariant by continuous deformation. The fact is that all open convex sets are equivalent, in particular equivalent to $\mathbf{R}^d$ itself. So for an algebraic topologist convex sets are of no interest. The proof is simple: build a continuous map $K \to \mathbf{R}^d$ by picking any $x \in K$ and stretching any non-infinite ray from x in K to the infinite associated ray in $\mathbf{R}^d$.

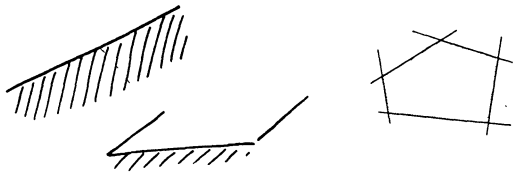


The following explicit examples of convex sets are basic:

(i) Each of the two half-spaces (open or closed) defined by an affine hyperplane (namely the set given by an equation like

$$\sum_{i=1}^{d} a_i x_i = b,$$

where the a_i's are not all 0) is convex. Half-spaces are the building blocks of convexity, since it is classical and not hard to prove that every closed convex set is the intersection of the closed half-spaces which contain it (see section 6). Moreover one can always achieve it with a denumerable family.



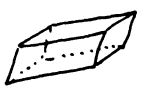
(ii) The full (closed) *ellipsoids*

$$\sum_{i=1}^{d} \frac{x_i^2}{a_i^2} \leq 1.$$

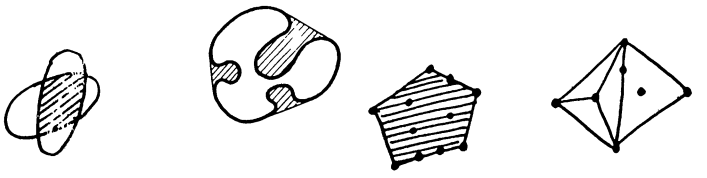
It is very important to remark that—when no Euclidean structure is imposed on $\mathbf{R}^d$—all ellipsoids are equivalent (affinely). They are equivalent, in particular, to the standard closed ball $\sum_{i=1}^{d} x_i^2 \leq 1$.



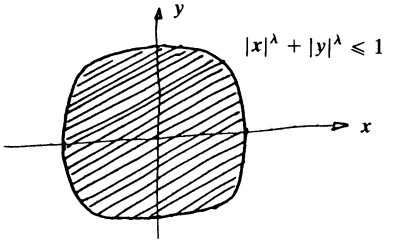
(iii) A closed *parallelepiped* is a set which can be written (after translation) as $\{(x_1, \ldots, x_d)\colon |x_i| \leq 1\ \forall i = 1, \ldots, d\}$ in a suitable basis of $\mathbf{R}^d$. They are all equivalent—affinely—to the standard *cube* $|x_i| \leq 1$ $(i = 1, \ldots, d)$ (the coordinates are the standard ones of $\mathbf{R}^d$). But as soon as metric considerations are imposed, then parallelepipeds can be very different.



(iv) Any intersection (not even necessarily countable) of convex sets is convex. Hence, it makes sense to speak of the smallest convex set containing a given set $A \subset \mathbf{R}^d$ and to call it the *convex hull* of A. It will be *denoted* by $\operatorname{conv}(A)$ in the sequel. *Polytopes* are the convex hulls of *finite* sets. Note that not all points are necessary; the really necessary ones are called the *vertices*, or extreme points, of the polytope.

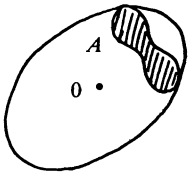


(v) The sets $|x|^\lambda + |y|^\lambda \leq 1$ in $\mathbf{R}^2$ are convex for every real number $\lambda \geq 1$. The proof is not obvious, but see section 4.



3. The John-Loewner ellipsoid; applications

We present here a result that is both simple—though nontrivial—and extremely powerful. It was found independently by F. John (studying mechanics) and C. Loewner (studying complex variable maps), both in the '40s. (The case of $d = 2$ goes back to Behrend in the '30s.) It says that given any bounded set A (with nonvoid interior) in $\mathbf{R}^d$, there exists one and *only* one ellipsoid E of minimal volume containing A (recall that ellipsoids are centered at the origin).



The volume of an ellipsoid $\Sigma_{i=1}^d (x_i^2/a_i^2) \leq 1$ is to be understood in the elementary sense, namely as the canonical measure on $\mathbf{R}^d$ (with the Euclidean standard metric). It is given by

$$\mathrm{Vol}\left(\sum_{i=1}^{d} \frac{x_i^2}{a_i^2} \leq 1 \right) = \left(\prod_{i=1}^{d} a_i \right) \cdot \mathrm{Vol}\left(\sum_{i=1}^{d} x_i^2 \leq 1 \right).$$

This will follow for example from the determinant change of variable rule in integration theory. We *denote* by $\beta(d)$ the volume of the standard ball $\Sigma_{i=1}^d x_i^2 \leq 1$. The value of this important function can be found in some calculus books (see [**6**])

namely

$$\beta(d = 2k) = \frac{\pi^k}{k!}$$

$$\beta(d = 2k + 1) = \frac{2^{k+1}\pi^k}{1 \cdot 3 \cdot 5 \cdots = (2k + 1)}$$

or in one shot

$$\beta(d) = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2} + 1\right)}$$

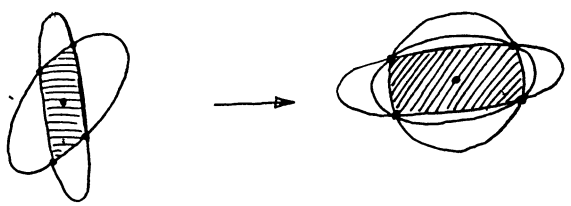
if you know the Γ-function.

If you have never thought about it, study the behavior of $\beta(d)$ as d increases. Compute in particular the first value of d for which $\beta(d) < 1$ (the volume of the cube of side equal to 1). An asymptotic evaluation of $\beta(d)$ is important (see section 6). From Stirling's formula one gets

$$\beta(d) \sim \text{constant} \cdot \left(\frac{2\pi e}{d}\right)^{(d+2)/2} \quad \text{as } d \to \infty.$$

It also permits one to compare $\beta(d)$ with the volume, equal to 2^d, of the cube circumscribed by the unit ball. See an interesting application at the end of section 5.A.

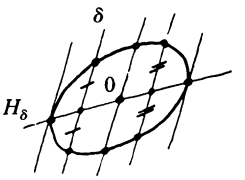
Let us come back to the John-Loewner assertion. Existence is an easy compactness argument. Just be careful to avoid the degeneracy of ellipsoids. This is guaranteed precisely by the non-void interior condition. To prove uniqueness argue by contradiction. Show that, given two distinct ellipsoids of the same volume, there exists a third one of smaller volume which contains their intersection. Simplify the computation by using the reduction of positive quadratic forms to simultaneous diagonal forms. Another proof is given in the next section.



The John-Loewner ellipsoid is widely used nowadays in the study of convex sets, for both pure and applied mathematics, see [**19**]. For our purpose let us mention three applications to various fields.

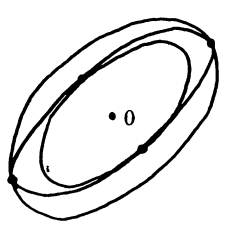
The first application is to the theory of quadrics (conics for the plane). An ellipsoid E has the property (called diametrical) of admitting an affine hyperplane symmetry for any direction δ of lines in $\mathbf{R}^d$. That is to say there exists a hyperplane H_δ associated to δ such that the hyperplane symmetry defined by the pair (δ, H_δ) leaves E invariant. Are the ellipsoids the only subsets A of $\mathbf{R}^d$ enjoying this property for any direction of line? The answer is easily seen to be yes if we use the

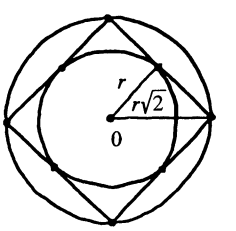
John-Loewner ellipsoid E of the set A. A bare-handed proof—you should try it—is already extremely involved (and not very enlightening) when $d = 2$. For $d = 3$, it was done by Bertrand and by Brunn.



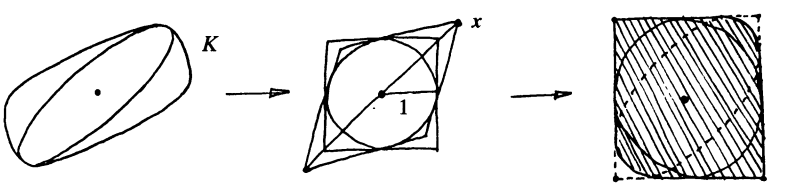
The second application is in geometric-group theory: it states that *every compact* subgroup G of the full linear group $GL(d;\mathbf{R})$ of $\mathbf{R}^d$ leaves invariant some Euclidean structure on $\mathbf{R}^d$. Take any point $x \neq 0$ in $\mathbf{R}^d$ and introduce the John ellipsoid of its orbit $G(x)$. This ellipsoid yields the desired quadratic form; the uniqueness is of course basic. Note more generally that the compact group we are studying can be taken in the full group $Aff(d;\mathbf{R})$ of all affine transformations of $\mathbf{R}^d$ (we permit translations). For by the Archimedes result of our introduction, the center of mass of the orbit $G(x)$ is invariant under G so we can take it to be the origin 0. Properly rephrased, the above existence result is group theoretical: two maximal compact subgroups G, G' of $GL(d;\mathbf{R})$ are necessarily conjugate, i.e., there exists g in $GL(d;\mathbf{R})$ such that $G' = gGg^{-1}$. This is actually a special case of a general result of Elie Cartan to the effect that this conjugacy of compact maximal subgroups is valid in any Lie group.

The third application is in fact what F. John had in mind. Call a *convex body* a non-void-interior compact set of $\mathbf{R}^d$. Then for every convex body K of $\mathbf{R}^d$ which is symmetric around the origin ($\forall x \in K$ then $-x \in K$) there exists an ellipsoid E such that $E \subset K \subset \lambda K$ and $\lambda \leq \sqrt{d}$. The bound is clearly optimal as shown by the cube.





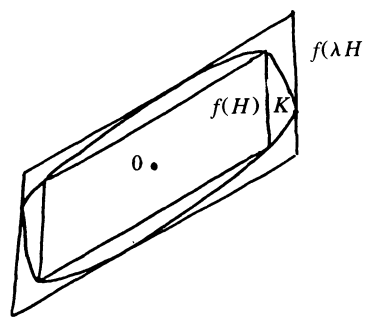
Again the proof works by contradiction. Using duality (see section 6 if necessary) introduce the ellipsoid of *maximal* volume contained in K and think of it (after a suitable linear transformation—remember (ii) in section 2) as the unit sphere in $\mathbf{R}^d$.



Assume there exists an $x \in K$ with $\|x\| > \sqrt{2}$. By the symmetry assumption K will contain the domain shaded above. Elementary calculus (again use a reduction to the standard cube) shows that the sphere inscribed in that cube is contained in K and has a larger volume than that of any inscribed ellipsoid.

It is worth mentioning the $\sqrt{d}$-squeezing property for two reasons.

The first is the *Banach-Mazur metric structure* on the set of all symmetric convex bodies in $\mathbf{R}^d$ which is defined as follows: for two convex bodies K, H (symmetric in the origin) call λ the smallest number for which there exists a linear transformation f of $\mathbf{R}^d$ such that $f(H) \subset K \subset f(\lambda H)$. Then $d(K, H) = \lambda$ is called the Banach-Mazur distance (it is in fact a so-called multiplicative distance and strictly speaking after dividing the set of convex bodies by the set of linear isomorphisms). John's result is now simply that the distance between any symmetric convex body and the unit ball is always $\leq \log\sqrt{d}$. It says that a Banach space structure on $\mathbf{R}^d$ is never too far from a Euclidean structure!



The second reason has to do with the very simple question: what is the greatest Banach-Mazur distance to the unit cube? This is a case (see the introduction) where the answer is not known today. The aim is to squeeze a symmetric convex body between two homothetic parallelepipeds as close together as possible. The value $\sqrt{d}$ is wrong. Szarek proved in 1987 that one needs at least $\sqrt{d}\log d$. The optimal bounds are still unknown. Two more comments. First, Szarek's construction of examples is not really explicit but based on probability theory. This technique is now widely used in convexity theory. The idea is that for a "general" convex body K the points of contact of K with the optimal cube are far from an orthonormal basis. On the contrary the angles are quite small. Second—a very general remark—for practical applications to theoretical harmonic analysis or to numerical analysis, it is more important to have asymptotic estimation when the dimension d goes to infinity than explicit values. In that direction, Szarek and Talagrand proved in 1988 that asymptotically the Banach-Mazur distance between the cube and any symmetric convex body is always of the order of $d^{7/8}$. The conjectured order is $\sqrt{d}\log d$ (up to some universal constant). See [**33**].

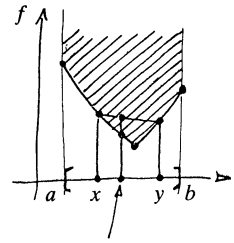
4. Convex functions; examples and applications

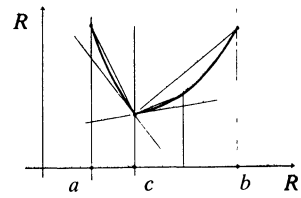
The simplest notion to start with is that of a convex real valued function defined on an interval—say closed $[a, b]$—of the real line:

$f\colon [a, b] \to \mathbf{R}$ is said to be convex (resp. strictly convex) if:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \text{ (resp., } <)$$

$$\forall a \leq x < y \leq b, \qquad 0 < \lambda < 1.$$





Otherwise stated: the graph of f is always below the chord segment joining every pair of its points. More simply: f convex is equivalent to requiring that the *epigraph* of f (namely the set $\{(x, y)\colon x \in [a, b],\ y \geq f(x)\} \subset \mathbf{R}^2$) is convex. A trivial induction yields

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i) \qquad \forall n\ \forall x_i \in [a, b]\ \forall \lambda_i \geq 0 \quad \text{with} \sum_{i=1}^{n} \lambda_i = 1.$$

Convex functions are necessarily quite regular: they everywhere admit a right and a left derivative which need not coincide, though the set at which they do not coincide can be at most countable. In particular they are almost everywhere (i.e., up to a set of measure 0) differentiable, in fact continuously differentiable. They also admit almost everywhere a second derivative $f'' \geq 0$. Most important is the converse: if f'' exists and is non-negative (resp., positive) everywhere then f is convex (resp. strictly convex). This easy result is tremendously powerful. Two standard and basic examples are

(i) $f(x) = -\log x$ on $[1, \infty[$ yields the inequality

$$\sum_{i=1}^{n} \lambda_i a_i \geq \prod_{i=1}^{n} a_i^{\lambda_i} \qquad \left(\lambda_i \geq 0,\ \sum_{i=1}^{n} \lambda_i = 1\right)$$

in particular

$$a_1 \cdots a_n \leq \left(\frac{a_1 + \cdots + a_n}{n}\right)^n.$$

This is easy for $n = 2$, but from $n = 3$ on it is not so obvious.

(ii) $f(x) = x^p$, $p \geq 1$. After tricky (but not deep) manipulations, one gets the Hölder inequality:

$$\sum_i x_i y_i \leq \left(\sum_i x_i^p\right)^{1/p} \left(\sum_i y_i^q\right)^{1/q} \quad \text{with } \frac{1}{p} + \frac{1}{q} = 1$$

and the Minkowski inequality:

$$\left(\sum_i |x_i + y_i|^p\right)^{1/p} \leq \left(\sum_i |x_i|^p\right)^{1/p} + \left(\sum_i |y_i|^p\right)^{1/p}.$$

The latter says that the set $|x|^\lambda + |y|^\lambda \leq 1$ is convex in $\mathbf{R}^2$. This is not obvious, even for $\lambda = 4$.

The notion of convexity can be extended to numerical functions defined on a subset K of $\mathbf{R}^d$:

$$f(lx + (1 - l)y) \leq lf(x) + (1 - l)f(y)$$

so that an obvious necessary condition for this condition to make sense is that K should be a convex subset of $\mathbf{R}^d$. General convex functions are also quite regular.

They always admit a differential almost everywhere, in fact they are almost everywhere of class C^1. More: almost everywhere of class C^2, i.e., they admit a second differential $f''(x)$: $\mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ which is there a positive quadratic form. The converse—analogous to the interval case—is also true.

Convex functions have two basic properties, widely used in theoretical and applied mathematics. Their maximum is attained only at the boundary of their domain of definition. A strictly convex function admits at most one minimum. John-Loewner ellipsoid uniqueness might also be proved using the latter property. Identify ellipsoids in $\mathbf{R}^d$ with quadratic forms (positive definite) and consider the set of all of them whose seat is in $\mathbf{R}^{d(d+1)/2}$. This is done, for example, with the map

$$\sum_{i,j} a_{ij}x_i x_j \to \underset{\|}{((a_{ij}))} \in \mathbf{R}^{d(d+1)/2}.$$
$$a_{ji}$$

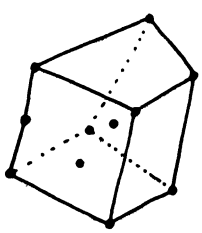
It is typical of how mathematicians build up successive levels of abstraction. The volume $\mathrm{Vol}(E(q))$ of the ellipsoid associated with q is equal to $\det(q)^{-1/2} \cdot \beta(d)$, where det stands for determinant. Then check that the function $q \to \det(q)^{-1/2}$ is strictly convex.

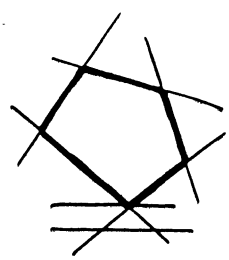
We will give ample consideration to the regularity of convex functions in section 7. A basic example is given in sections 5 and 6, the Brunn-Minkowski inequality.

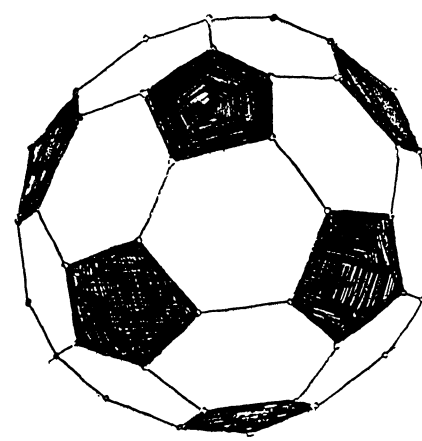
5. Polytopes: Four "elementary" problems

Recall that a *polytope* is the convex hull of finite points in some $\mathbf{R}^d$; if $d = 2$ we call it a polygon (convex), and if $d = 3$ a polyhedron (convex). Its *vertices* are the really necessary points; any that are not useful should be thrown out. It is not hard to see (in section 6, for example) that a polytope is also a finite intersection of half-spaces. Conversely, one should add compactness. The *faces* ($(d-1)$-dimensional faces, to be exact) are the intersections of the polytope with the *really necessary* hyperplanes which define it. Such a face is a polytope in its hyperplane.

By induction one defines the *i-faces* of a polytope ($i = 0, 1, \ldots, d-1$). The $(d-1)$-faces are the faces, the 0-faces are the vertices, the 1-faces are the *edges*. In dimension 3 this exhausts them.





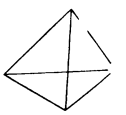


"the most popular polyhedron"

Examples of polytopes are parallelepipeds. The simplest are the simplices. A *simplex* is a polytope generated by the minimum number of points in $\mathbf{R}^d$ for it to be a convex body (in particular not contained in some hyperplane). This number is $d + 1$. All simplices are affinely the same. Thus, not surprisingly, simplices are of basic utility (e.g., the simplex method of Dantzig in linear analysis).



$d = 2$: triangle



$d = 3$: tetrahedron

Because of their simple definition and the everyday look they have about them, it is natural to expect that:

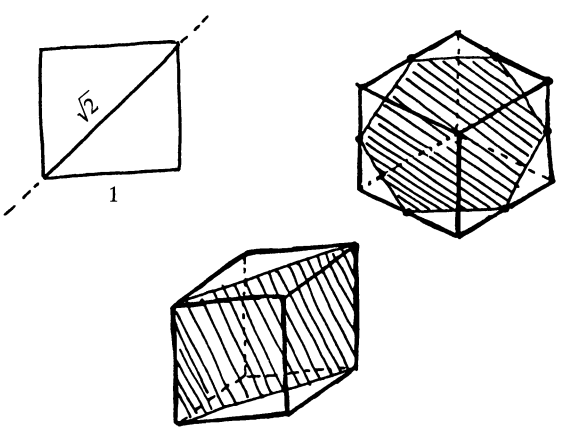
a) everything about polytopes has been known for a long time

b) everything about polytopes is easy to prove.

In the light of our introduction, the reader will already have guessed that both statements are false. We will illustrate this with four topics.

A. Hyperplane sections of the cube

Consider the unit cube $C = [-\frac{1}{2}, \frac{1}{2}]^d$ in $\mathbf{R}^d$ and cut it by hyperplanes. Which hyperplane cuts C with maximum volume ($(d-1)$-dimensional measure, that is to say its volume for standard measure in the hyperplane where it stands, for the natural Euclidean structure)?



For $d = 2$ the answer is $\sqrt{2}$, attained by either of the two diagonals. For $d = 3$ the smart reader will guess that the most wonderful section of the cube is the regular hexagon obtained by cutting it through the origin by a hyperplane orthogonal to a diagonal. But he will be wrong because its area is equal to $3\sqrt{3}/4$, whereas the section through two opposite diagonals of parallel faces has area $\sqrt{2} > 3\sqrt{3}/4$. It was conjectured for quite a while that $\sqrt{2}$ is the optimum (clearly attained) for any d. This was proved by K. Ball only in 1986, see [**2**]. The proof is by no means elementary. It is based on probability theory and, at the root of it lies the Fourier

transform, applied twice to get estimates of the volume of sections by hyperplanes. One then gets the following beautiful formula for the section of C by the hyperplane H given by the equation $\sum_{i=1}^{d} a_i x_i = 0 (\sum_{i=1}^{d} a_i^2 = 1)$:

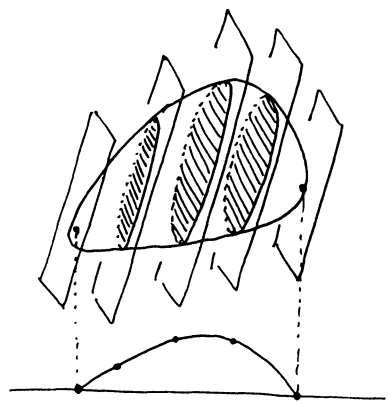
$$\text{volume}(C \cap H) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin a_1 t}{a_1 t} \cdots \frac{\sin a_d t}{a_d t} \cdot dt.$$

Then one finishes the proof with Hölder inequality (under the integral form of the finite form given in section 4) and the fact that

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \left| \frac{\sin t}{t} \right|^p dt \leq \frac{\sqrt{2}}{\sqrt{p}} \quad \text{if } p \geq 2$$

and equality only if $p = 2$. This latter inequality is quite subtle to prove.

An interesting consequence of Ball's result is the very easy disproof of a (quite intuitive) conjecture of Herbert Busemann: if two convex bodies K and H (say symmetric in the origin) are such that the volume of the section of K by any hyperplane P through the origin is always smaller than the volume of the section of H by P, then the volume of K is smaller than the volume of H. This is trivially true when $d = 2$; it is open for $d = 3, \ldots, 7$. A quite involved counterexample was given by Larman and Rogers in 1975 for $d = 12$. But from Ball's result it follows immediately that the conjecture is false for any $d \geq 10$ when K is the unit cube in $\mathbf{R}^d$ and H the sphere in $\mathbf{R}^d$ of radius so adjusted that its volume is equal to one.



Here is a small remark in the same spirit. Cut any symmetric convex body by parallel hyperplanes. You expect the volumes of these sections to be a monotone function, starting at a minimum when the hyperplane just touches the body and reaching a maximum through the origin. This is easy to prove with the Brunn-Minkowski inequality (see section 6), but however obvious it may seem, no elementary proof is known. Further: the Brunn-Minkowski inequality is equivalent to the fact that the $1/(d-1)$ power of the volume of the parallel hyperplane sections of any convex body (symmetric or not) in $\mathbf{R}^d$ is a *concave* function. It is only in dimension two that this is completely elementary.

Note also that, although apparently obvious, it is not easy to prove that any hyperplane section through the center of a unit cube has a volume greater than or equal to one. It was proven by Vaaler only in 1979!

B. Combinatorics

Let us *denote* by f_i the number of the i-dimensional faces of a polytope Γ of $\mathbf{R}^d$. Here i runs from 0 to $d-1$. Are there necessary (and sufficient) conditions for a given sequence $(f_0, f_1, \ldots, f_{d-1})$ of integers to be that of some polytope in $\mathbf{R}^d$? The case $d=2$ is trivial: $f_0 = f_1$ is both necessary and sufficient. When $d=3$ Euler found in 1750 the famous necessary condition $f_0 - f_1 + f_2 = 2$ (which seems also to have been found—though none too explicitly—by Descartes). It is not hard to extend this relation to any d in the form $f_0 - f_1 + \cdots + (-1)^{d-1} f_{d-1} = (-1)^d$.

It was only in the 1920's that Steinitz found the set of necessary and sufficient conditions when $d=3$. It reads:

$$f_0 - f_1 + f_2 = 2 \quad \text{and} \quad 4 < f_0 < 2f_2 - 4 \leq 4f_0 - 12.$$

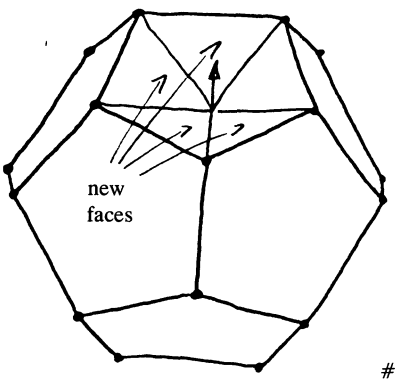
The inequalities were known to Descartes and Euler as necessary ones. For $d \geq 4$ such a set is unknown. But in the 1980's two basic results were obtained. They concern generic polytopes, namely the so-called *simplicial polytopes*. These are polytopes (in $\mathbf{R}^d$) all of whose $(d-1)$-faces are $((d-1)-1)$ simplices. Equivalently, their vertices are in general position, that is to say that there are never more than d of them belonging to the one affine hyperplane. They are completely *flexible* in the sense that every vertex can be moved a little bit when all the others are kept fixed.

For simplicial polytopes a necessary and sufficient set of conditions is known, though only since 1980. It is quite complicated. The set of conjectured necessary and sufficient conditions was discovered by McMullen. The way the proofs go is most interesting. The sufficiency—due to Billera and Lee—is obtained by exhibiting the desired polytope from the very involved commutative algebra construction of a Cohen-Macaulay ring. The proof of the necessity, due to Stanley, is even more surprising. It is obtained (following an initial idea of Demazure) by associating with a convex polytope a complex projective algebraic variety and applying to it the so-called "hard Lefschetz theorem." This story again illustrates the need to use complex number geometry even in order to solve real number problems. See [29], [32].

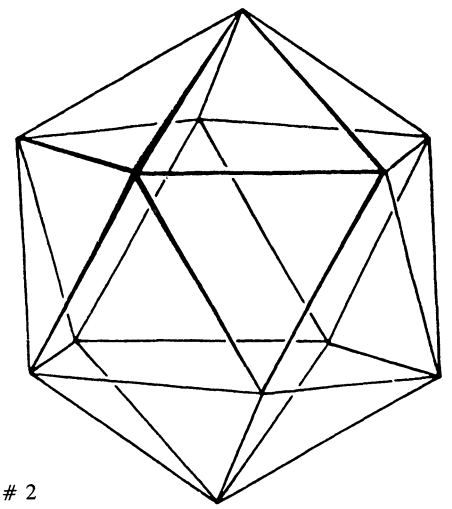
C. Can you move a polytope? Can you put it into a computer?

Consider a polyhedron in $\mathbf{R}^3$ and try to move (slightly) its vertices under the condition that the polyhedron remains of the same combinatorial type, i.e. it keeps the same number of faces, each face keeping the same number of vertices. More algebraically, during the move you want to respect the linear relation which exists between the vertices of the initial polytope. Of course you discard the trivial solutions such as an affine linear transformation of the whole space. The end of the last section shows that, for a *simplicial* polyhedron (i.e., one of which all the faces are triangles) the movability is maximum, every vertex can be moved (slightly) independently of the others. The problem begins to be interesting when one has faces with four or more vertices. In the figures below it will be seen that you cannot move only one vertex, like a, because this will force new faces into existence. For the dodecahedron, one way is to move the plane of a face slightly. But this is possible only because a dodecahedron is *cosimplicial* (see section 7), that is that when considered as intersection of half-spaces, only three planes meet at one vertex. But this trick does not work for polyhedron No. 3. It can be done by

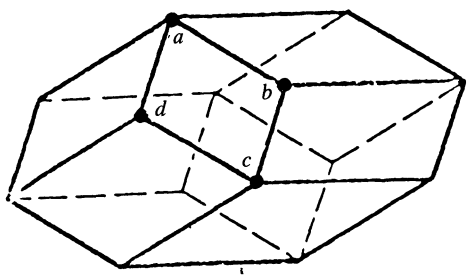
turning the plane of the face *abcd* around *a* and *c*. Then the general situation seems a very complicated problem for a general polyhedron. In 1920 Steinitz proved a more general theorem about the general polyhedron, to the effect that some graphs can always be realized in $\mathbf{R}^3$ by a convex polyhedron. As a corollary, any convex polyhedron has sufficient freedom of movement to be approximated as well as required by a polyhedron all of whose vertices have rational number coordinates (call it a *rational* polyhedron).



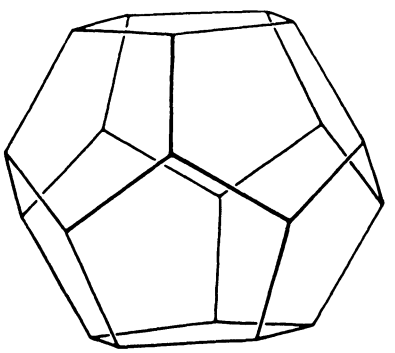
1



2



3

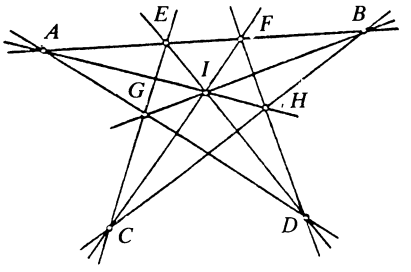


4

This is important because, once a coordinate system is fixed, a computer can recognize only rational polyhedrons when they are entered by giving the three coordinates of the vertices. This means that, for example, the affine linear relations between vertices have to be truly respected, not only approximated. To end with dimension 3, let us mention that we still do not know how to compute the effective degree of freedom when the combinatorial type is given.

How about higher dimensions? The situation here is dramatically different because in 1967 Perles found a polytope in $\mathbf{R}^8$ with twelve vertices which cannot be approximated by a rational one. The idea is roughly as follows. The arrangement

of faces will force, in the diagram below,



the dotted points to be on lines as indicated. Now an elementary computation in the projective plane (using, if you know it, the notion of cross-ratio) shows that whatever the Euclidean metric is, we should have

$$\frac{EA}{EB} \cdot \frac{FB}{FA} \quad \text{equal to} \quad \frac{3-\sqrt{5}}{2} \quad \text{or} \quad \frac{3+\sqrt{5}}{2}.$$

But neither of these numbers is rational! See [**22**].

In 1987 Sturmfels found a polytope in $\mathbf{R}^6$ with the non-rational approximation property. Whether this can be done in $\mathbf{R}^5$ or even $\mathbf{R}^4$ is still an open question. The above shows, in particular, that you cannot implement in a computer every diagram exactly respecting the alignments. For the aficionados, a computer expert will tell you that there are other "theoretical" ways to enter such a diagram into a computer.

A comment on *rigidity* is in order. For the above, the movability was combinatorial, not metric. It was permitted to change the lengths of the edges (and the congruence of the faces). If, moreover, we insist, when moving a polyhedron in $\mathbf{R}^3$, that all the faces remain *equal* polygons (this is automatically true if all faces are triangles and if the edges are kept at fixed lengths) then Cauchy proved rigidity back in 1812, namely that only a global Euclidean displacement can achieve it. In fact there was a small gap in Cauchy's proof that Hadamard and Steinitz filled in independently. Note that it is false for plane polygons as soon as they have at least four vertices.

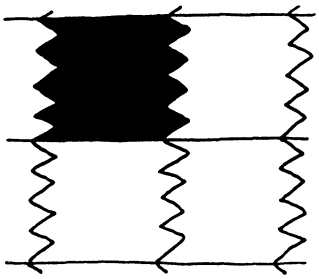
With a little bit of spherical geometry and induction on dimension, Cauchy's theorem implies rigidity in any dimension larger than three. Cauchy's proof is very subtle. See, for example [**6**].

D. Fillers

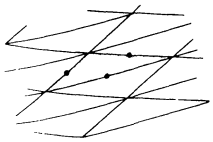
Call a polygon P a *filler* of the *Euclidean* plane $\mathbf{R}^2$ if the whole plane can be filled up ("tiled") with congruent (equal) copies of P; the copies should moreover meet edge to edge.

We first want to find the shape of all possible fillers. Any triangle and any quadrilateral will do. Then there are examples with pentagons and hexagons. The hexagonal case is completely understood, the pentagonal one not yet ([**24**]). Geometers (in fact crystallographers) have known since the end of the 19th century that no *convex* k-gon with $k \geq 7$ can be a filler. Convexity is of course required. It would seem that there is no completely elementary proof. The known proofs use

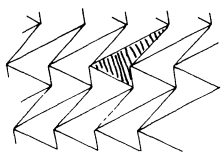
not only Euler's formula but also a "going to infinity" argument. For fillers of the plane see Thurston's text in the present issue, and the basic reference [**24**].

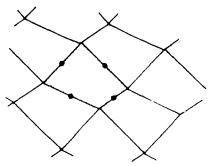


no limit on the number of edges!

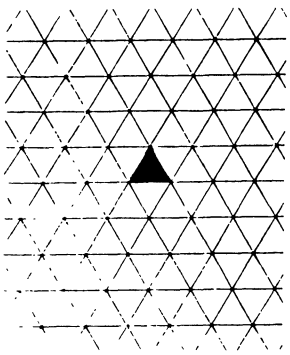


for triangles use symmetries

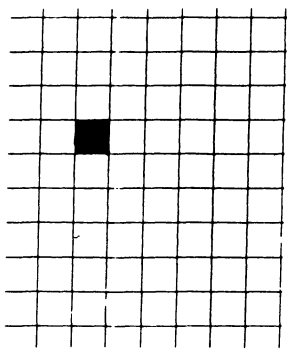




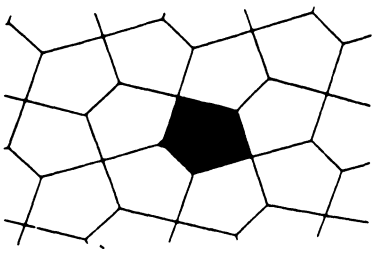
same for quadrilaterals



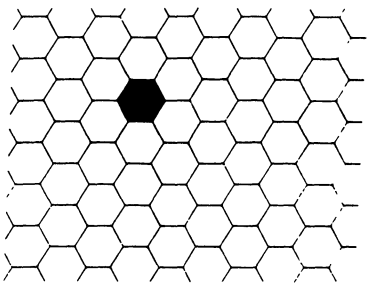
3



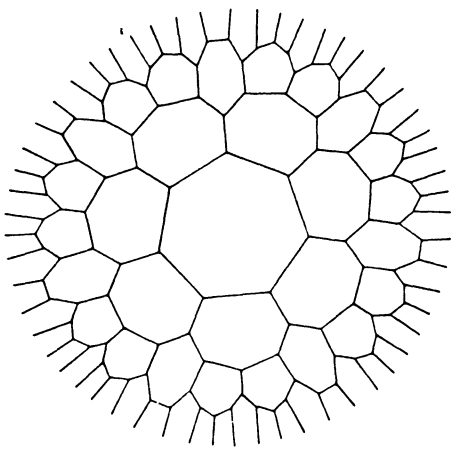
4



5



6



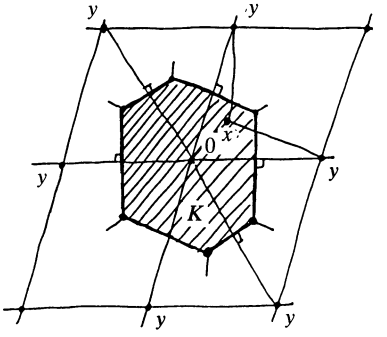
7 but not congruent tiles

We now look at $\mathbf{R}^3$. Call a *filler* a polytope P such that the whole space $\mathbf{R}^3$ can be filled with congruent copies of P, still matching face to face.

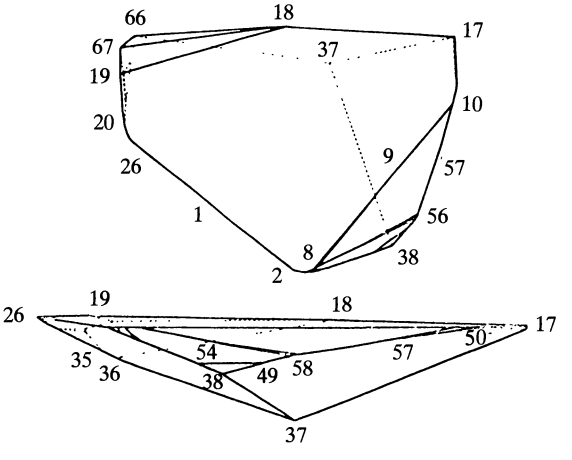
Here two surprises await us. The first is that there do exist extremely complicated fillers. Engel found one in 1980 with 38 faces and 70 vertices. Some faces are extremely small. This is surprising at first: any filler is surrounded by 38 others and some of the contacts are very tiny. The construction used a computer because of these small faces. It is based on the classical and important notion of what is known as the Dirichlet-Voronoï domains. We start with a lattice Λ in $\mathbf{R}^3$ and define P by $P = \{x \in \mathbf{R}^3 : \forall \lambda \in \Lambda : d(x,0) \leq d(x,\lambda)\}$. Then, by the very construction, the polytopes $\{P + \lambda : \lambda \in \Lambda\}$ fill $\mathbf{R}^3$. Now Engel's example is obtained by extending the Dirichlet-Voronoï technique as follows. We consider discrete groups G of isometries of $\mathbf{R}^3$ which contain a lattice of translations, but in fact G can also contain rotations and be much bigger than only a lattice group. Then the Engel filler P is of the type

$$P = \{x \in \mathbf{R}^3 : \forall g \in G : d(x,0) \leq d(x, g(0))\}.$$

This P is in general called the *fundamental domain* of the group G.



But there is another surprise: there is no known bound for the number of vertices or faces of an $\mathbf{R}^3$ filler when it is not obtained by a discrete group of isometries techniques. We do not even know that it *is* bounded! See [**23**].

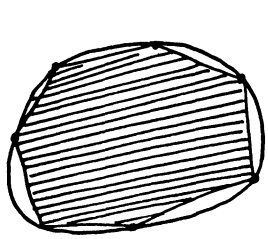


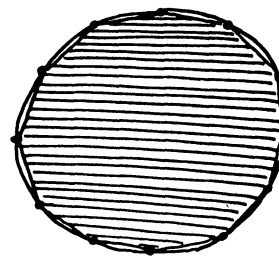
Two views of a 38-faced polytope with 70 vertices, discovered by Engel (1980)

6. Two algebraic operations on the set of all convex bodies: Duality and addition

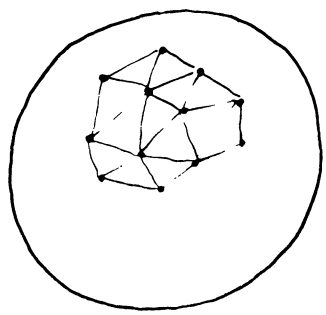
The typical evolution of mathematical understanding consists in reaching higher and higher levels of abstraction, by the building of structures. In order to study individual convex bodies, we are going, here and in the next section, to study the set of all of them (in a given dimension, say d) and to introduce algebraic and topological structures on this set.

The duality notion, a very useful one in mathematics (think of the Fourier transform analysis, homology—cohomology, ...), will be introduced through the expedient of a practical problem, that of calculating the volume of a convex body by computer. Looking at polygons for curves in the plane it looks simple:





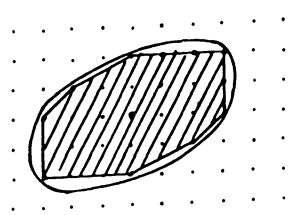
Archimedes on π



But in large dimensions we will meet surprises.

A. Evaluating the volume of a convex body by computer

Let us assume that a convex body K is described, in the computer, by a membership oracle: input a point $\vec{x}$, and the oracle announces whether or not $\vec{x} \in K$.



Lovász found, in 1985, a polynomial time algorithm that will give for $\text{vol}(K)$ a lower bound *denoted* by $\underline{\text{vol}}(K)$. The crucial problem lies in how good that

approximation is. Barany and Füredi, in 1987, show it to be a complete disaster. They prove that for an input which is polynomial in d, say of $n = d^a$ points, one cannot *in general* do better than

$$\frac{\text{vol}(K)}{\overline{\text{vol}}(K)} \leq \left(\frac{d}{2ac \log d}\right)^{d/2}.$$

For the proof we consider the simplest convex body, the unit ball B^d. It can be proved that the maximal volume of a polytope with $n = d^a$ points contained in B^d, if denoted by $V(d, n)$, obeys

$$\frac{V(d,n)}{\text{vol}(B^d)} \leq \left(\frac{2ac \log d}{d}\right)^{d/2}.$$

The idea is roughly as follows. By Caratheodory's theorem (see section 8.E) every point of the convex hull of $n = d^a$ points belongs to a simplex of $\mathbf{R}^d$. If the points are within distance 1 from the origin, subtle estimates can be obtained for the volume of such a simplex from some of its k-faces. Then one will have at most $\binom{n}{k}$ such situations to fill out the polytope and a $1/k!$ factor coming from the volume of a k-dimensional simplex. The subtlety now lies in a smart choice for k. The answer is to take k as the integral part of $d/2 \log n$. Then, because of the formulas given at the beginning of section 3 for the volume of the unit ball B^d and $\binom{n}{k} = n!/k!(n-k)!$, one concludes, after computations using the Stirling formula for evaluating $p!$ when p is large, which asserts that

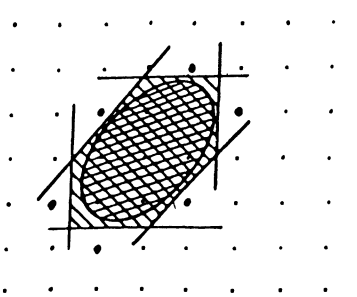
$$\frac{p!}{\sqrt{2p\pi}\left(\frac{p}{e}\right)^p} \xrightarrow[p\to\infty]{} 1.$$

A comment: ellipsoids are really the *worst* convex sets for the above inequality. This was proved in 1951 by Macbeath. When the number of points n is given, for convex bodies K in a given $\mathbf{R}^d$, the bound

$$\frac{\text{Sup(Volume of a polytope contained in } K \text{ with } n \text{ vertices)}}{\text{vol}(K)}$$

is minimum for ellipsoids.

There is still one hope. Try to estimate $\text{vol}(K)$ from below by $\underline{\text{vol}}(K)$ with a membership oracle, and try to get an upper bound $\overline{\text{vol}}(K)$ with a membership and a *separation* oracle together, which are easy to implement. When the computer answers $x \notin K$ it will at the same time give you a half-space containing K:



Still according to Lovász, there is also a polynomial algorithm giving $\underline{\mathrm{vol}}(K) \geq \mathrm{vol}(K)$. Moreover,

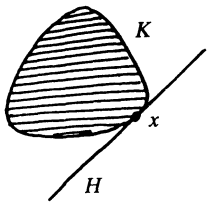
$$\frac{\overline{\mathrm{vol}}(K)}{\underline{\mathrm{vol}}(K)} \leq d^d.$$

This is bad for large ds. But one might hope to be smart enough in choosing points to get a good

$$\frac{\overline{\mathrm{vol}}(K)}{\underline{\mathrm{vol}}(K)}.$$

This amounts to being able to play with points and half-spaces (or, say, affine hyperplanes). This is precisely what the classical duality in Euclidean spaces achieves for us. The final result will be given in C below (after this duality has been explained in B).

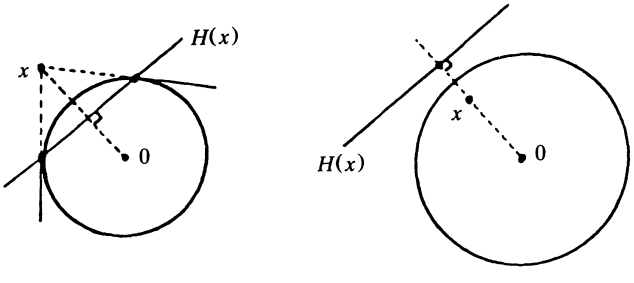
But before we end with a great classic of convexity, let us recall that Minkowski showed just before 1900 that any convex body admits at least one *supporting* hyperplane H (affine here) for any point x of its boundary:



namely, K lies entirely in one of the two closed half-spaces determined by H. In most books this is called the Hahn-Banach theorem because Hahn and Banach proved it, much later, in the infinite-dimensional case. This basic result for the study of convex sets shows in particular what we announced in section 2: a closed convex set is the intersection of the closed half-spaces which contain it.

B. The Euclidean duality

We consider $\mathbf{R}^d$ with its standard Euclidean structure. The duality between points of $\mathbf{R}^d$ and its affine hyperplanes is geometrically defined as follows:



To $x \neq 0$ associate the line $0x$ and call the *dual* of x the hyperplane $H(x)$ which is orthogonal to $0x$ and cut it at the point x' such that $0x \cdot 0x' = 1$. Conversely, x,

clearly unique for any hyperplane H not containing 0, is called the *dual* (the *pole*) of H. Check for yourself in the plane what the dual of the intersection of two lines is in $\mathbf{R}^2$. Algebraically, things are much simpler:

$$H(x) = \{y \in \mathbf{R}^d : (x|y) = 1\}$$

where $(\cdot|\cdot)$ denotes the usual scalar product.

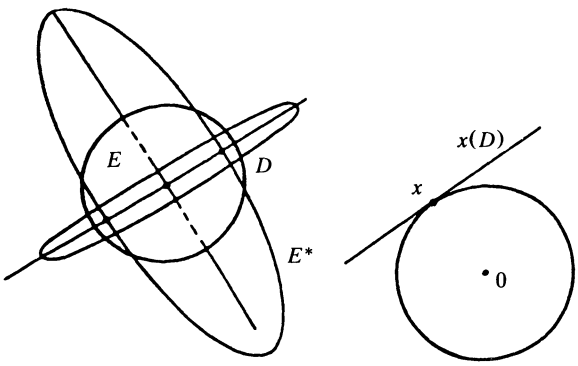
The *polar* (*reciprocal*) convex body K^* of a given convex body K is either the convex hull of the poles of the supporting hyperplanes of H or the intersection of the half spaces bounded by the hyperplanes which are the dual of the points of K (of course you need to use the boundary points of K). This duality is excellent if K is a convex body that contains 0 in its interior. From now on only such K's will be considered. Then $(K^*)^* = K$.

Examples are:

(i) the unit ball B^d is its own dual

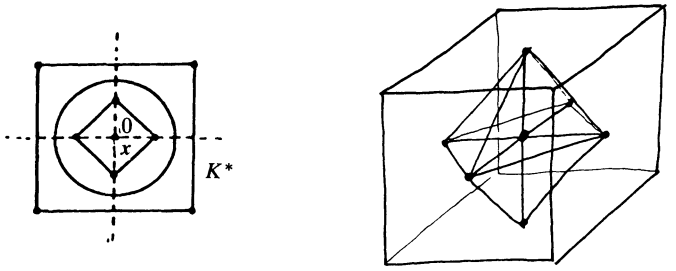
(ii) ellipsoids have ellipsoids as duals; check that then

$$\mathrm{vol}(E)\mathrm{vol}(E^*) = \left(\mathrm{vol}(B^d)\right)^2 = \beta^2(d)$$



(iii) the dual of a cube is called a *cross polytope*. Precisely the dual of $[-1, 1]^d = K$ is $K^* =$ convex hull of the point $\pm e_i$ (where $\{e_i\}$ is the standard base). Note that a cube has 2^d vertices, the cross polytope only $2d$. Note here

$$\mathrm{vol}(K)\mathrm{vol}(K^*) = 2^d \frac{2^d}{d!} = \frac{4^d}{d!}.$$



(iv) more generally the dual of a polytope is a polytope, the duality exchanging vertices of one with faces of the other.

C. Back to estimating volumes: the invariant $\mathrm{vol}(K)\,\mathrm{vol}(K^*)$

Here we work with convex sets K symmetric in the origin. The picture and the way Lovász obtains $\mathrm{vol}(K)$ and $\overline{\mathrm{vol}}(K)$ show that if we want to compute $\mathrm{vol}(B^d)$ by his method, we have to estimate

$$\boxtimes \qquad \frac{\overline{\mathrm{vol}}(B^d)}{\mathrm{vol}(B^d)} \geq \frac{\mathrm{vol}(K^*)}{\mathrm{vol}(K)},$$

where K is one of the polytopes contained in B^d as in the figure in subsection A. We rewrite the right-hand quotient of $\boxtimes$ as

$$\mathrm{vol}(K)\mathrm{vol}(K^*)\left(\frac{\mathrm{vol}(B^d)}{\mathrm{vol}(K)}\right)^2 \frac{1}{(\mathrm{vol}(B^d))^2}.$$

The last two terms are estimated above. The plan is to work with a smart K, that is to approximate B^d by K such that $\mathrm{vol}(K)\mathrm{vol}(K^*)$ is very small.

Note that $\mathrm{vol}(K)\mathrm{vol}(K^*)$ is an invariant of the linear shape of K only because linear transformations will affect $\mathrm{vol}(K)$ by $\det f$ when it affects $\mathrm{vol}(K^*)$ by $(\det f)^{-1}$. It was conjectured by Mahler that for any d and any K:

$$\beta^2(d) \geq \mathrm{vol}(K)\mathrm{vol}(K^*) \geq \frac{4^d}{d!},$$

the left inequality being characteristic of ellipsoids and the right one of cubes or cross polytopes.

Assume Mahler's conjecture on the right side. Then the hopes mentioned above are dashed. Because then always

$$\frac{\mathrm{vol}(K^*)}{\mathrm{vol}(K)} \geq \left(\frac{d}{ae\log d}\right)^d$$

using the above results.

As to Mahler's double conjecture, the situation today is the following. The left bound is true and characterizes the ellipsoid. The bound was proved by Blaschke for $d = 2, 3$ and by L. Santalo (with some restrictions) for any d. The fact that equality characterizes the ellipsoid for any d was settled only by Saint Raymond in 1981. The exact right-hand bound is unknown today even for $d = 3$. It is known only for $d = 2$ (try it yourself to feel the difficulty there already is for $d = 2$). In 1985, J. Bourgain and V. Milman showed that the order of magnitude conjectured by Mahler is the right one. There exists $c > 0$

$$\mathrm{vol}(K)\mathrm{vol}(K^*) \geq \frac{c^d}{d!} \quad \forall K \ \forall d.$$

So the error on computing volumes will have to be as bad as

$$c\left(\frac{d}{\log d}\right)^d.$$

Thus "it is impossible to compute the volume with a decent error in polynomial time." In other words a computation of volume applicable to any convex body should be of another type.

The Bourgain-Milman proof is quite involved. It uses probability theory and fine estimates, see [8], [5].

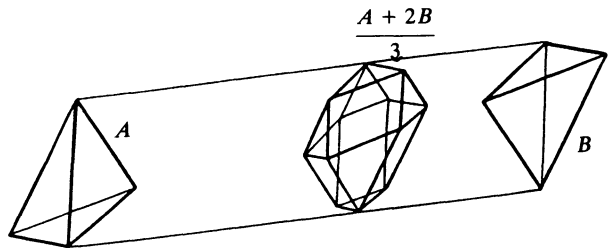
Note that the Mahler conjecture is useful in a completely different context, that of simultaneous approximations of real numbers by rationals, see [17], page 31.

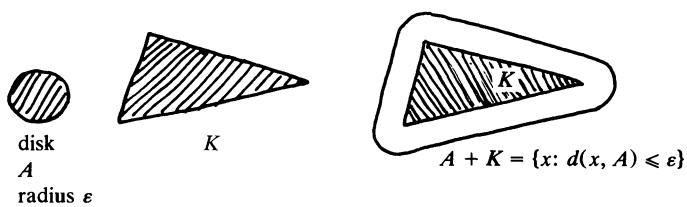
D. The Minkowski addition

For K, H convex sets of $\mathbf{R}^d$, the set defined as

$$K + H = \{x + y: x \in K, \quad y \in H\}$$

is still convex.





disk
A
radius ε

Note that if you change the origin in $\mathbf{R}^d$ the new $K + H$ will differ from the old one by only a translation, so it can be considered as well-defined in shape. The notion is then affine. More generally, $\lambda K + \mu H = \{\lambda x + \mu y: x \in K, y \in H\}$ is still convex for any $\lambda, \mu > 0$. The picture below shows

$$\frac{A + 2B}{3} = \frac{1}{3}A + \frac{2}{3}B$$

for two simplices in $\mathbf{R}^3$ and also that the sum of some K with a ball of radius ε is nothing but the set of points within a distance ε of K. We have no time to elaborate on this. Suffice it to note that it is the starting point of one of the first rigorous proofs of the standard isoperimetric inequality in $\mathbf{R}^d$ which states that among all convex bodies of given volume, balls are those with the smallest boundary volume (here $(d-1)$-volume which is length when $d = 2$ and area when $d = 3$), see section 8. Let us mention here the basic Brunn-Minkowski inequality

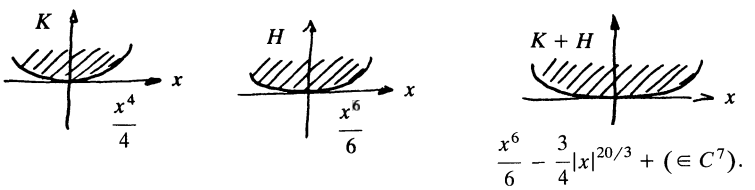
$$(\mathrm{vol}(A + B))^{1/d} \geq (\mathrm{vol}(A))^{1/d} + (\mathrm{vol}(B))^{1/d}.$$

This was essentially discovered by Brunn in 1888. The reader will easily see that this is exactly equivalent to the claim that the $(d-1)^{-1}$th power of the $(d-1)$-dimensional volume of the sections of a given convex body in $\mathbf{R}^d$ is a *concave* function when the cutting hyperplane moves from the left to the right: see section 5.A.

In both of the above figures, the Minkowski addition has a *regularization effect*; in both cases $K + H$ has a smoother boundary than K, H. This is naïve for the general case but was discovered only in 1988 by Kieselman. Even when K and H have the most regular boundary possible—namely, C^∞, or real analytic or even more: algebraic—the boundary of $K + H$ can be non-smooth. The results are of local nature. The negative one consists simply of the epigraphs K, H (see section 4) of $x^4/4$ and $x^6/6$ for which

$$K + H \text{ is the epigraph of } \frac{x^6}{6} - \frac{3}{4}|x|^{20/3} + \text{something of class } C^7.$$

So it is only six times differentiable. But it is the worst possible. The positive result is: if the boundary of K and the boundary of H are of class C^∞, then the boundary of $K + H$ is always of class $C^{20/3} = C^{6+(2/3)}$. See [**26**].



So the Minkowski addition has no regularizing effect when regularity means *local* smoothness (differentiability properties). But it does have a *global* regularizing effect when regularity means being close to ellipsoids. It was indeed proved by Vitali Milman that for any convex body in $\mathbf{R}^d$ there exist two affine isomorphisms A and B of $\mathbf{R}^d$ such that if one sets $T = K + A(K)$ then the dual T^* of T has the advantage that $T^* + B(T^*)$ has a Banach-Mazur distance to the sphere always bounded by a number independent of the dimension d.

7. Topology on the set of all convex bodies: intuition is dangerous

A. Topology

We are going to see with the help of topology that intuition can again be misleading. Namely that if any convex body has a boundary which is almost everywhere C^1 and C^2 differentiable, in the set of all convex bodies almost every convex body has a boundary which is never everywhere C^2. For even worse, see below. Topology is needed to give a precise meaning to "almost every convex body."

For the sake of simplicity we work in the set $\mathscr{K}$ of all convex bodies symmetric in the origin of a given $\mathbf{R}^d$. We recall that these are convex sets which are also compact and with a non-void interior. We specify $\mathscr{K}$ in $\mathbf{R}^d$ only when needed. Topology on the convex body set is best given by the *Banach-Mazur metric*.

We have already seen in section 3 the Banach-Mazur distance on $\mathscr{K}$. It has the advantage of being linear-invariant. But, *stricto sensu*, it is a distance in fact only when convex sets which differ by some linear transformation are identified. A non-linear-invariant distance is the Hausdorff one. Fix on $\mathbf{R}^d$ some Euclidean metric. Then $d(K, H)$ is defined by the smallest ε such that $K \subset H + \varepsilon B^d$ *and* $H \subset K + \varepsilon B^d$, i.e., K is within distance ε from H *and* conversely. It is easy to

prove that the Hausdorff distance gives the same topology on $\mathscr{K}$ as the Banach-Mazur one (just beware of the quotient to be done).

Remark that no reasonable measure on $\mathscr{K}$ is known. To define the analogue of "almost every" without measure but when one has a topology, we use Baire's idea: *negligible* sets will be by definition those which are a denumerable union of sets with void interior (these latter are also called *meager*).

B. Probable regularity of the boundary of a convex set

Consider a convex body K and its boundary ∂K. We want to study the regularity of this set which is a curve when $d = 2$ and a surface when $d = 3$. In general a hypersurface. This is a local question. In particular, taking $x \in \partial K$ and a supporting hyperplane H at x we can locally consider K as the epigraph of a convex numerical function f: $U \to \mathbf{R}$ where U is a neighborhood of x in H. Then ∂K will locally coincide with the graph of f.

We can apply the results quoted in section 4. So if we build the totality of ∂K with such graphs we know that ∂K is almost everywhere (this for the natural measure on ∂K obtained from the Lebesgue measures on the H's through the f's) a differentiable manifold of class C^2 (in particular almost everywhere of class C^1). The question we address now is: what are the chances for ∂K to be everywhere of class C^1 when we pick K at random in $\mathscr{K}$. A nice answer was provided by Klee in 1959. He proved that, up to a negligible set in K, ∂K is always C^1 everywhere, hence a nice C^1-submanifold of $\mathbf{R}^d$ (of codimension one).

In contrast to this the C^2 situation is dramatically different: Gruber in 1977 proved that the set of K in $\mathscr{K}$ whose ∂K is almost everywhere C^2 is negligible. Worse: Zamfirescu in 1980 proved that, except in a negligible set, ∂K has horrible curvature properties: curvature does not exist on a non-denumerable set and where it does exist it is equal to zero! The reason is that that is exactly how the curvature of a polytope behaves and it so happens that in $\mathscr{K}$ polytopes are dense. See [**20**], [**37**].

8. Brief survey of other important topics in convexity

For lack of space we cannot treat these in detail but only briefly mention what they are about and provide some references.

A. Inequalities

The most important one is Brunn-Minkowski's mention in section 6. It yields the classical isoperimetric inequality in $\mathbf{R}^d$. For other proofs, other inequalities, especially the *mixed volume* inequalities, which are quite deep and hard to prove, see [**11**], [**27**].

B. Almost spherical sections, concentration phenomena, applications

The isoperimetric inequality on spheres was used by Paul Lévy back in the '30s to prove concentration properties for functions. It was used by Milman in 1971 to prove the inspiring theorem of Dvoretsky (1961) that any convex body admits almost ellipsoidal sections for suitable codimension. This was the starting point of a whole series of recent works in the same vein. Its importance lies, besides its strong geometrical appeal, in its application (through the so-called asymptotic methods) to Banach spaces of infinite dimension. For this see [**28**].

C. Convexity in analysis

Recently discoveries have been made of various functions which are log-concave, like $\mathrm{vol}^{1/d}$ in the Brunn-Minkowski inequality. It seems to have started with the important Brascamp-Lieb inequality which solved, in particular, the long-standing conjecture: the level lines of the first eigenfunction of the Dirichlet problem of a convex domain are convex hypersurfaces. In fact the complete heat kernel has a convexity property. See [3], [9].

D. Packing and covering

These two notions are very important. They have been connected, since Minkowski, with number geometry, rational approximations. But also with error detecting codes. Interesting references are [13], [17].

They are also connected with notions of entropy and infinite-dimensional Banach geometry: see [28].

E. Caratheodory, Helly, Radon

Three elementary (and very visualizable) statements are:

the *Caratheodory theorem*: the convex hull of a set in $\mathbf{R}^d$ is generated by the positive barycenters of its $(d+1)$-uples

the *Helly theorem*: a family of compact convex sets has a non-void intersection as soon as any $d+1$ of its sets has a non-void intersection

the *Radon theorem*: let T be a subset of at least $d+2$ points in $\mathbf{R}^d$. Then there always exists a disjoint decomposition $T = T_1 \cup T_2$ such that $\mathrm{conv}(T_1) \cap \mathrm{conv}(T_2) = \varnothing$.

None of these theorems are very hard to prove. Nor are they too elementary. The very interesting fact is that any one of them easily implies the two others. Moreover they have an enormous number of applications and are linked a great deal with combinatorics. See, for example [15], [18].

F. Convexity in other spaces

Convexity can be defined in more general spaces than the Euclidean ones. In particular, it is widely used nowadays in Riemannian Geometry. See [4], [31].

G. Last but not least: The moment map

This map was introduced very recently. The fact that it is convex in certain cases is very powerful and turns out to be a surprising link between mechanics, Kähler geometry, Lie group theory and eigenvalues of matrices. See [1], [25].

REFERENCES

Besides the items mentioned within the text and listed thereafter, the reader will find systematic (although never, of course, exhaustive) treatment of convexity in [6], [7], [10], [12], [16], [22], [27], [35].

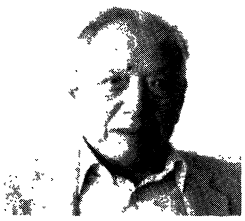
1. M. F. Atiyah, Angular Momentum, Convex Polyhedra and Algebraic Geometry, *Proc. Edinburgh Math. Soc.* 1983 (26) 121–138.
2. K. Ball, Cube slicing in $\mathbf{R}^n$, *Proceed. A.M.S.*, 97 (1986) 465–473.
3. _____, Logarithmically concave functions and sections of convex sets in $\mathbf{R}^n$ *Studia Math.*, 88 (1988) 69–84.

4. W. Ballmann, M. Gromov, and V. Schroeder, Manifolds of Negative Curvature, Birkhäuser, 1985.
5. I. Barany and Z. Füredi, Computing the volume is difficult, *Discrete and Computational Geometry*, 2 (1987) 319–326.
6. M. Berger, Geometry I and II, Springer, 1987.
7. T. Bonnesen and W. Fenchel, Theorie der konvexen Körper, Springer (reprint), 1974.
8. J. Bourgain and V. D. Milman, New volume ratio properties for convex symmetric bodies in $\mathbf{R}^n$, *Inventiones Math.*, 88 (1987) 319–341.
9. H. J. Brascamp and E. H. Lieb, Best constants in Young's inequality, its converse and its generalization to more than three functions, *Advances in Math.*, 20 (1976) 151–173.
10. A. Bronsted, An Introduction to Convex Polytopes, Springer, 1983.
11. J. D. Burago and V. A. Zalgaller, Geometric Inequalities, Springer, 1988.
12. H. Busemann, Convex Surfaces, Interscience, 1958.
13. J. H. Conway and N. J. A. Sloane, Sphere Packings, Lattices and Groups, Springer, 1988.
14. J. Coolidge, Algebraic Plane Curves, Dover, 1959.
15. L. Danzer, B. Grünbaum, and V. Klee, Helly's theorem and its relatives, *AMS Proc. Sympos.*, 7, 1963.
16. H. G. Eggleston, Convexity, Cambridge University Press, 1958.
17. P. Erdös, P. M. Gruber, and J. Hammer, Lattice Points, Pitman Monographs and Surveys in Pure and Applied Mathematics, No. 39, John Wiley, 1989.
18. W. J. Ellison, Waring's problem, this MONTHLY, 78 (1971) 10–36.
19. M. Grötschel, L. Løvász, and A. Schrijver, Geometric Algorithms and Combinatorial Optimization, Springer, 1988.
20. P. M. Gruber, Aspects of convexity and its applications, *Expositiones Math.*, 2 (1984) 47–83.
21. P. M. Gruber and J. M. Wills (editors), Convexity and its Applications, Birkhäuser, 1983.
22. B. Grünbaum, Convex Polytopes, Interscience, 1967 (or U.M.I. Out-of-Print Books on Demand, 1989).
23. B. Grünbaum and G. C. Shephard, Tilings with Congruent Tiles, *Bull. AMS*, 3 (1980) 951–973.
24. _____, Tilings and Patterns, Freeman, 1987.
25. V. Guillemin and S. Sternberg, Convexity properties of the moment mapping, *Invent. Math.*, 67 (1982) 491–513.
26. C. O. Kiselman, Smoothness of vector sums of plane convex sets, *Math. Scand.*, 60 (1987) 239–252.
27. K. Leichtweiss, Konvexe Mengen, Springer, 1980.
28. V. D. Milman and G. Schechtman, Asymptotic Theory of Finite Dimensional Normed Spaces, Springer Lecture Notes in Math., No 1200, 1986.
29. T. Oda, Convex Bodies and Algebraic Geometry, Springer, 1988.
30. C. A. Rogers, Packing and Covering, Cambridge University Press, 1964.
31. K. Shiohama, Topology of complete and non-compact manifolds, *in* Advanced Studies in Pure Mathematics, No 3: Geometry of Geodesics and Related Topics, Kinokuniya and North Holland, 1984.
32. R. P. Stanley, Combinatorics and Commutative Algebra, Birkhäuser, 1983.
33. S. J. Szarek and M. Talagrand, An isomorphic version of the Sauer-Shelah lemma and the Banach-Mazur distance to the cube, preprint IHES, 1988.
34. J. Tölke and J. M. Wills (editors), Contributions to Geometry, Birkhäuser, 1979.
35. F. A. Valentine, Convex Sets, McGraw-Hill, 1964.
36. R. J. Walker, Algebraic Curves, Princeton Univ. Press, 1950.
37. T. Zamfirescu, Nearly all convex bodies are smooth and strictly convex, *Monatshefte für Math.*, 103 (1987) 57–62.
38. _____, Curvature properties of typical convex surfaces, *Pacific J. Math.*, 131 (1988) 191–207.

What Is Geometry?

SHIING-SHEN CHERN*, *MSRI, Berkeley, CA 94720*

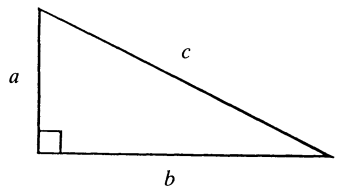
SHIING-SHEN CHERN. Born October 26, 1911 in Kashing, Chekiang Province, China, Dr. Chern was educated at Nankai University, Tsinghua University, and the University of Hamburg, where he received his doctorate in 1936. He has been Professor at UC Berkeley since 1960, Emeritus since 1979, and is now Director Emeritus of MSRI in Berkeley as well as Director of the Nankai Institute of Mathematics in Tianjin, China. He is a member of the National Academy of Sciences and of numerous other Academies, domestic and foreign. He has received the National Medal of Science of the USA, and many other prizes, distinctions and honors.



To avoid misunderstanding I will not give a definition of geometry as in the customary mathematical treatment of a topic. I will only try to discuss its major historical developments.

1. Geometry as a logical system; Euclid. Euclid's "Elements of Geometry" (ca. 300 B.C.) is one of the great achievements of the human mind. It makes geometry into a deductive science and the geometrical phenomena as the logical conclusions of a system of axioms and postulates. The content is not restricted to geometry as we now understand the term. Its main geometrical results are:

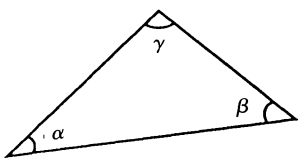
a) Pythagoras' Theorem.



$$c^2 = a^2 + b^2$$

FIG. 1

b) Angle-sum of a triangle.



$$\alpha + \beta + \gamma = 180°$$

FIG. 2

The result b) is derived using the fifth, or the last, postulate, which says: "And that, if a straight line falling on two straight lines make the angles, internal and on

*Work done under partial support of NSF grant DMS-87-01609

the same side, less than two right angles, the two straight lines, being produced indefinitely, meet on the side on which are the angles less than two right angles."

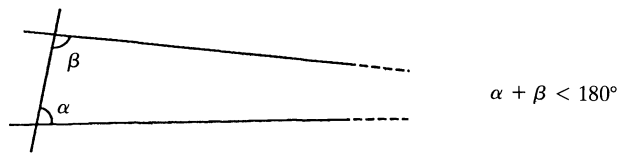


FIG. 3

Euclid realized that the parallel postulate was not as transparent as his other axioms and postulates. Efforts were made to prove it as a consequence. Their failure led to the discovery of non-Euclidean geometry by C. F. Gauss, John Bolyai, and N. I. Lobachevski in the early 19th century.

The "Elements" treated rectilinear figures and the circle. The last three of its thirteen Books were devoted to solid geometry.

2. Coordinatization of space; Descartes. The introduction of coordinates by Descartes (1596–1650) was a revolution in geometry. In the plane it can be described by the following figure:

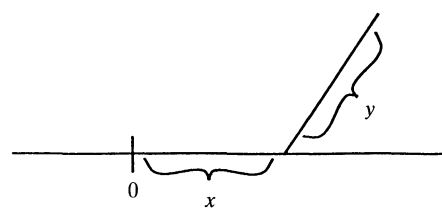


FIG. 4

where the role of the two coordinates x, y is not symmetric. Descartes' work was published in 1637 as an appendix, entitled "La géométrie", to his famous book on philosophy [**6**]. At about the same time Fermat (1601–1665) also found the concept of coordinates and used them to treat successfully geometric problems by algebraic methods. But Fermat's work was published only posthumously [**7**].

One immediate consequence was the study of curves defined by arbitrary equations

$$F(x, y) = 0, \tag{1}$$

thus enlarging the scope of the figures.

Fermat went on to introduce some of the fundamental concepts of the calculus, such as the tangent line and the maxima and minima.

From two dimensions one goes to n dimensions, and to an infinite number of dimensions. In these spaces one studies loci defined by arbitrary systems of equations. Thus a great vista was opened, and geometry and algebra became inseparable.

A mystery is the role of differentiation. The analytic method is most effective when the functions involved are smooth. Hence I wish to quote a philosophical question posed by Clifford Taubes [**15**]: Do humans really take derivatives? Can they tell the difference?

Coordinate geometry paved the way to applications to physics. An example was Newton's derivation of the Kepler laws from his law of gravitation. Kepler's first law says that the planetary orbits are ellipses with the sun as their common focus. The proof was possible only after an analytic theory of conics had been established.

3. Space based on the group concept; Klein's Erlanger Programm. Works on geometry led to the development of projective geometry, among whose founders were: J. V. Poncelet (1788–1867), A. F. Möbius (1790–1868), M. Chasles (1793–1880), and J. Steiner (1796–1863). Projective geometry studies the geometrical properties arising from the linear subspaces of a space and the transformations generated by projections and sections. Other geometries resulted, the most notable ones being affine geometry and conformal geometry.

In 1872 Felix Klein formulated his Erlanger Programm [**1**], [**11**], which defines geometry as the study of the properties of a space that are invariant under a group of transformations. Thus there is a geometry corresponding to every group of transformations acting on a space. The basic notion is "group" and the notion of a space is now greatly expanded. In a certain sense the group of projective collineations is the most encompassing group and projective geometry occupies a dominant position.

The most important application of the Erlanger Programm was the treatment of non-Euclidean geometry by the so-called Cayley-Klein projective metric [**12**]. The hyperbolic space can be identified with the interior of a hypersphere and the non-euclidean motions with the group of projective collineations leaving invariant the hypersphere. The same group may appear as a group of transformations in different spaces. As a result the same algebraic argument could give entirely different geometric theorems. For example, everybody knows that the three medians of a triangle meet in a point. By using Study's dual numbers this translates into the following theorem of J. Petersen and F. Morley: Let *ABCDEF* be a skew hexagon such that consecutive sides are perpendicular. The three common perpendiculars of the pairs of opposite sides AB, DE; BC, EF; CD, FA have a common perpendicular. See [**13**].

Sophus Lie founded a theory of general transformation groups, which became a fundamental tool of all geometry.

4. Localization of geometry; Gauss and Riemann. In his monograph on surface theory published in 1827 [**8**], Gauss (1777–1855) developed the geometry on a surface based on its fundamental form. This was generalized by B. Riemann (1826–1866) to n dimensions in his Habilitationschrift in 1854 [**14**]. Riemannian geometry is the geometry based on the quadratic differential form

$$ds^2 = \sum g_{ik}(u)\, du^i\, du^k, \qquad g_{ik} = g_{ki}, \qquad 1 \le i, k \le n \tag{2}$$

in the space of the coordinates $u^1, \ldots, u^n$, where the form is positive definite, or at least non-degenerate. Given ds^2, one can define the arc length of a curve, the

angle between two intersecting curves, the volume of a domain, and other geometrical concepts.

The main characteristic of this geometry is that it is local: it is valid in a neighborhood of the u-space. Because of this feature it fits well with field theory in physics. Einstein's general theory of relativity interprets the physical universe as a four-dimensional Lorentzian space (with a ds^2 of signature $+++-$) satisfying the field equations

$$R_{ik} - \frac{1}{2}g_{ik}R = 8\pi\kappa T_{ik}, \tag{3}$$

where R_{ik} is the Ricci curvature tensor, R is the scalar curvature, κ is a constant, and T_{ik} is the energy-stress tensor.

It is soon observed that most properties of Riemannian geometry derive from its Levi-Civita parallelism, an infinitesimal transport of the tangent spaces. In other words, Riemannian geometry studies the tangent bundle of a Riemannian space with the Levi-Civita connection.

5. Globalization; topology. Riemannian geometry and its generalizations in differential geometry are local in character. It seems a mystery to me that we do need a whole space to piece the neighborhoods together. This is achieved by topology. The notion of a differentiable manifold is one of the most sophisticated concepts in mathematics. The idea was clear to Riemann. The first mathematical formulation of a topological manifold was made by D. Hilbert in 1902 [**10**], [**17**]. Hermann Weyl identified the Riemann surfaces with one-dimensional complex manifolds and used it as the central theme of his epoch-making book *Die Idee der Riemannschen Fläche* [**16**]. On the topological side "neighborhood" became the basic concept in Hausdorff's topology [**9**].

Hassler Whitney saw the merit of establishing an imbedding theorem on differentiable manifolds (1936), thus beginning the serious study of differential topology. That derivatives play a role in topology came as a shock when J. Milnor discovered the exotic differentiable structures on the seven-dimensional sphere (1956). By studying the Yang-Mills equations on a four-dimensional manifold, S. Donaldson found in 1983 a remarkable theorem on the intersection-form, which led to the existence of an infinite number of differentiable structures on R^4.

With the foundation of differentiable manifolds laid, geometrical structures can now be defined on them, such as the Riemannian structure, the complex structure, the conformal structure, the projective structure based on a system of paths, etc. Tools are developed for their treatment, of which the most important are the exterior differential calculus and the tensor analysis.

A fundamental notion is "curvature," in its different forms. Its simplest manifestation is the circle in plane Euclidean geometry. It could also be the force of a physical system or the strength of a gravitational or electro-magnetic field. In mathematical terms it measures the non-commutativity of covariant differentiation.

It is remarkable that suitable algebraic combinations of curvature give topological invariants. To illustrate this we wish to state the Gauss-Bonnet theorem. Let D be a domain with a sectionally smooth boundary on a two-dimensional Riemannian

manifold. Then the Gauss-Bonnet theorem is the formula

$$\sum(\pi - \alpha) + \int_{\partial D} k_g \, ds + \int_D \int K \, dA = 2\pi\chi(D), \tag{4}$$

where the first term is the sum of the exterior angles at the corners, the second term is the integral of the geodesic curvature along the sides, the third term is the integral of the Gaussian curvature over D, and $\chi(D)$ is the Euler characteristic of D. For a rectilinear triangle in the Euclidean plane this is the theorem on the angle-sum stated in §1. For higher dimensions we will only give, for the sake of simplicity, the theorem for a compact oriented Riemannian manifold M of dimension $2n$ without boundary. Let R_{ijkl} be the Riemann-Christoffel tensor and let

$$\Omega_{ij} = \sum R_{ijkl} \, du^k \wedge du^l \tag{5}$$

be the "curvature form". Let

$$Pf = \sum \epsilon_{i_1 \cdots i_{2n}} \Omega_{i_1 i_2} \wedge \cdots \wedge \Omega_{i_{2n-1} i_{2n}} \tag{6}$$

be the pfaffian, where $\epsilon_{i_1 \cdots i_{2n}}$ is $+1$ or -1 according as its indices form an even or odd permutation of $1, \ldots, 2n$, and is otherwise zero, and the sum is extended over all indices from 1 to $2n$. Then the Gauss-Bonnet theorem says

$$(-1)^n \frac{1}{2^{2n}\pi^n n!} \int_M Pf = \chi(M), \tag{7}$$

where $\chi(M)$ is the Euler-Poincaré characteristic of M.

6. Connections in a fiber bundle; Elie Cartan. A notion which includes both Klein's homogeneous spaces and Riemann's local geometry is Cartan's generalized spaces (espaces généralisés). In modern terms it is called "a connection in a fiber bundle." It is a straightforward generalization of the Levi-Civita parallelism, which is a connection in the tangent bundle of a Riemannian manifold. In general, we have a fiber bundle π: $E \to M$, whose fibers $\pi^{-1}(x)$, $x \in M$, are homogeneous spaces acted on by a Lie group G. A connection is an infinitesimal transport of the fibers compatible with the group action by G.

I wish to illustrate this more precisely in the case of a complex vector bundle, where the fibers are complex vector spaces C_q of dimension q and $G = GL(q; \mathbb{C})$ [4]. The importance of complex numbers in geometry is a mystery to me. It is well organized and complete. One manifestation is the simple behaviour of the group $GL(q; \mathbb{C})$: its maximal compact subgroup $U(q)$ has no torsion and has as Weyl group the group of all permutations on q letters.

We shall call a frame an ordered set of linearly independent vectors $e_1, \ldots, e_q \in \pi^{-1}(x)$, $x \in M$. In a neighborhood U where a frame field $e_1(x), \ldots, e_q(x)$, $x \in U$, is defined, a connection is given by the infinitesimal displacement

$$De_\alpha = \sum \omega_\alpha^\beta e_\beta, \qquad 1 \le \alpha, \beta \le q, \tag{8}$$

where ω_α^β are linear differential forms in U. We call ω_α^β the connection forms and the matrix

$$\omega = (\omega_\alpha^\beta) \tag{9}$$

the connection matrix. Under a change of the frame field

$$e'_\alpha = \sum a_\alpha^\beta e_\beta, \qquad A = (a_\alpha^\beta), \tag{10}$$

the connection matrix is changed as follows:

$$\omega' A = dA + A\omega. \tag{11}$$

We introduce the curvature matrix

$$\Omega = d\omega - \omega \wedge \omega, \tag{12}$$

which is a matrix of exterior two-forms. By exterior differentiation of (11) we get

$$\Omega' = A\Omega A^{-1}. \tag{13}$$

It follows that the exterior polynomial

$$\det\left(I + \frac{i}{2\pi}\Omega\right) = 1 + c_1(\Omega) + \cdots + c_q(\Omega), \tag{14}$$

in which $c_\alpha(\Omega)$ is a 2α-form, is independent of the choice of the frame field and is hence globally defined in M. Moreover, each c_α is closed, i.e.,

$$dc_\alpha = 0. \tag{15}$$

The form $c_\alpha(\Omega)$ has been called the αth Chern form of the connection and its cohomology class $\{c_\alpha(\Omega)\}$ in the sense of de Rham cohomology is an element of the cohomology group $H^{2\alpha}(M; \mathbb{Z})$ and is called the αth Chern class of the bundle E. These characteristic classes are the simplest and most fundamental global invariants of a complex vector bundle. They have the advantage of possessing a local representation, by curvature.

As in the Gauss-Bonnet formula such a representation is of great importance, because the forms $c_\alpha(\Omega)$ themselves have a geometrical significance. Moreover, let π': $P \to M$ be the bundle of frames of the complex vector bundle. Then the pull-back $\pi'^* c_\alpha$ becomes a derived form, i.e.,

$$\pi'^* c_\alpha = dTc_\alpha, \tag{16}$$

where Tc_α, a form of degree $2\alpha - 1$ in P, is uniquely determined by certain properties. This operation is called transgression and Tc_α have been called the Chern-Simons forms [**5**]. These forms have played a role in three-dimensional topology and in recent works of E. Witten on quantum field theory [**20**].

This theory can be developed for any fiber bundle; see [**3**]. The above provides the geometrical basis of gauge field theory in physics. Here M is a four-dimensional Lorentzian manifold, so that the Hodge $*$-operator is defined, and we define the codifferential

$$\delta = *d*. \tag{17}$$

There is a discrepancy of terminology and notation, as given by the following table:

mathematics	physics
connection ω	gauge potential A
curvature Ω	strength F

Maxwell's theory is based on a $U(1)$-bundle over M, and his field equations can be written

$$dA = F, \quad \delta F = J, \tag{18}$$

where J is the current vector. Actually, Maxwell wrote the first equation as

$$dF = 0, \tag{19}$$

which is a consequence. For most applications (19) is sufficient. But a critical study of an experiment proposed by Boehm and Aharanov and performed by Chambers shows that (18) are the correct equations [**21**]. A generalization of (18) to an $SU(2)$ bundle over M gives the Yang-Mills equations

$$DA = F, \quad \delta F = J. \tag{20}$$

It is indeed remarkable that developments in geometry have been consistently parallel to those in physics.

7. An application to biology. So far the most far-reaching applications of geometry are to physics, from which it is indeed inseparable. I wish to mention an application to biology, namely, to the structures of DNA molecules. This is known to be a "double helix", which geometrically means a pair of closed curves. Their geometrical invariants will clearly be of significance in biology. The following three are most important: 1) The linking number introduced by Gauss; 2) the total twist, which is essentially the integral of the torsion; 3) the writhing number.

James White proved that between these invariants there is the relation [**18**]

$$Lk = Tw + Wr. \tag{21}$$

This formula is of fundamental importance in molecular biology.

8. Conclusion. Contemporary geometry is thus a far cry from Euclid. To summarize, I would like to consider the following as the major developments in the history of geometry:

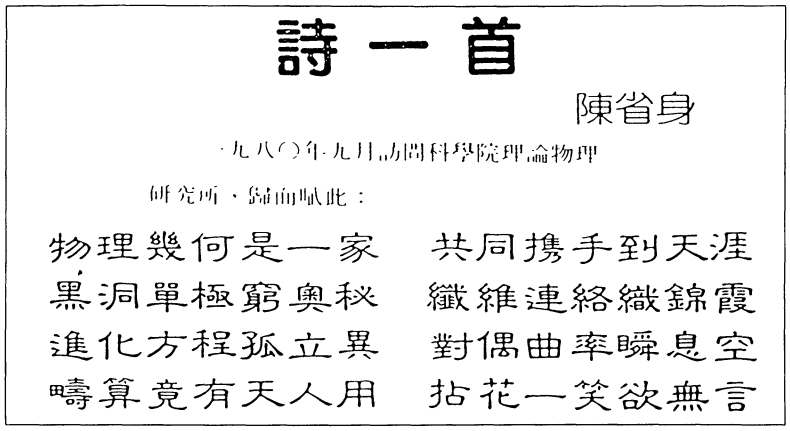
1) Axioms (Euclid);
2) Coordinates (Descartes, Fermat);
3) Calculus (Newton, Leibniz);
4) Groups (Klein, Lie);
5) Manifolds (Riemann);
6) Fiber bundles (Elie Cartan, Whitney).

A property is geometric, if it does not deal directly with numbers or if it happens on a manifold, where the coordinates themselves have no meaning. Going to several variables, algebra and analysis have a tendency to be involved with geometry.

This story is clearly one-sided and incomplete, representing only my personal viewpoint, and my limitations. It is clear that the story will not end here. Recent developments in theoretical physics, such as geometric quantum field theory, string theory, etc, are pushing for a much more general definition of geometry [**19**].

It is satisfying to note that so far almost all the sophisticated notions introduced in geometry have been found useful.

Finally, I wish to call attention to an early paper of mine [**2**], which could be read as a companion to this one.


詩一首

陳省身

一九八〇年九月訪問科學院理論物理

研究所，歸而賦此：

物理幾何是一家　共同携手到天涯
黑洞單極窮奧秘　纖維連絡織錦霞
進化方程孤立異　對偶曲率瞬息空
疇算竟有天人用　拈花一笑欲無言

REFERENCES

1. G. Birkhoff and M. K. Bennett, Felix Klein and His "Erlanger Programm", History and Philosophy of Modern Mathematics (W. Aspray and P. Kitcher, editors), Univ. of Minn. Press, 1988, 145–176.
2. S. Chern, From triangles to manifolds, this Monthly, 86 (1979) 339–349.
3. S. Chern, Complex Manifolds without Potential Theory, 2nd edition, Springer 1979.
4. S. Chern, Vector bundles with a connection, Studies in Global Differential Geometry, Math. Asso. Amer. Studies no. 27, (1989) 1–26.
5. S. Chern and J. Simons, Some cohomology classes in principal fiber bundles and their application to Riemannian geometry, Proc. Nat. Acad. Sci. USA, 68 (1971), 791–794; or, characteristic forms and geometrical invariants, *Annals of Math*, 99 (1974) 48–69.
6. René Descartes, Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences, 1637.
7. Pierre de Fermat, Oeuvres, edited by Paul Tannery and Charles Henry, Gauthier-Villars, Paris, 1891–1912.
8. C. F. Gauss, Disquisitiones generales circa superficies curvas, 1827; Ges. Werke, 4.
9. F. Hausdorff, Grundzüge der Mengenlehre, Leipzig 1914; dritte Auflage, Dover, N.Y. 1944; English translation, Chelsea, N.Y. 1957.
10. D. Hilbert, Uber die Grundlagen der Geometrie, Göttinger Nachrichten, 1902, 233–241.
11. F. Klein, Vergleichende Betrachtungen über neuere geometrische Forschungen, Math. Annalen 43 (1893), 63–100 or Ges. Abh I (1921), 460–497.
12. _____, Vorlesungen über nicht-euklidische Geometrie, Springer, 1928.
13. _____, Höhere Geometrie, Springer, 1926, p. 314.
14. B. Riemann, Über die Hypothesen welche der Geometrie zu Grunde liegen, Habilitationschrift 1854; Gött Abh 13, 1868; Ges. Werke 1892.
15. Clifford H. Taubes, Morse theory and monopoles; topology in longe range forces, Progress in Gauge Field Theory, Cargese 1983, 563–587, NATO Adv. Sci. Inst, Ser B, physics 115, Plenum New York-London, 1984.
16. H. Weyl, Die Idee der Riemannschen Fläche, Leipzig, 1913; 3 te Auflage, verändert, Leipzig, 1955.
17. _____, Riemanns geometrische Ideen, ihre Auswirkung und ihre Verknüpfung mit der Gruppentheorie, Springer, 1988.
18. James H. White, Self-linking and the Gauss integral in higher dimensions, Amer. J. Math., 91 (1969) 693–728.
19. E. Witten, Physics and geometry, Proc. Int. Cong. of Math. Berkeley 1986, Amer. Math. Soc., 1987, Vol. 1, 267–303.
20. _____, Quantum field theory and the Jones polynomial, Braid Group, Knot Group, and Statistical Mechanics, (C. N. Yang and M. L. Ke editors), World Scientific, 1989, 239–329.
21. C. N. Yang, Magnetic monopoles, fiber bundles, and gauge fields, Annals of New York Academy of Sciences, 294 (1977), 86–97.

A Whirlwind Tour of Computational Geometry

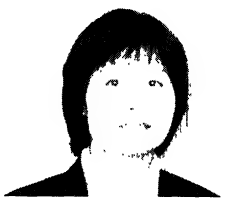
RON GRAHAM, *AT & T Bell Laboratories, Murray Hill, NJ 07974*

FRANCES YAO, *Xerox PARC, Palo Alto, CA 94304*

RON GRAHAM, well known to our readers, has won numerous distinctions in mathematics. He is a member of the National Academy of Sciences, has received the Pólya prize in combinatorics, is an editor of numerous journals, is a trustee of the AMS, etc. His research is in combinatorics, graph theory, number theory and algorithms.



FRANCES YAO is now Principal Scientist and Manager of the Theoretical Computer Science Area at Xerox Palo Alto Research Center. She received her Ph.D. in mathematics from MIT in 1973 and has taught at several universities, including Stanford (1976–1979). Her research interests span theoretical computer science, with emphasis on computational geometry.



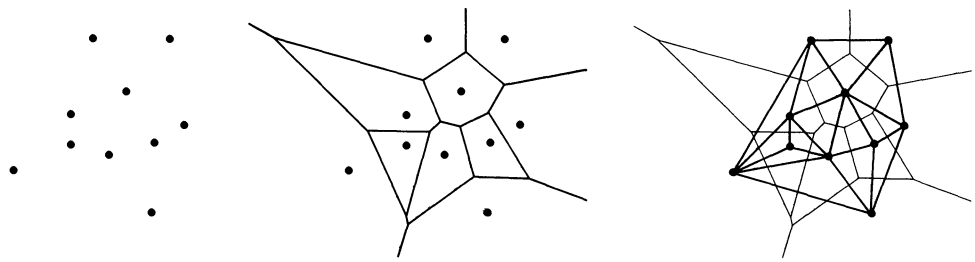
Computational geometry is concerned with finding efficient *algorithms*, or computational procedures, for solving a wide spectrum of geometric problems. Such problems can arise from computer graphics, robotics and motion planning, computer-aided design and manufacturing, or one of the many other applied areas where geometry comes into play. In spite of the practical flavor of these problems, results from classical and modern geometry quite often can play key roles in the design of efficient methods for solving them. Indeed, Euclidean geometry could be thought of as perhaps the earliest systematic study of algorithms: the ruler-and-compass constructions in Euclidean geometry are just algorithms based on a well defined set of elementary operations. In this venerable subject also occur some of the first provably unsolvable problems, such as squaring the circle, trisecting the angle and doubling the cube, much in the same spirit as the more recent negative solution of Hilbert's tenth problem (implying the nonexistence of an algorithm for solving general diophantine equations) [**9**] and the celebrated (but still unresolved) P vs. NP question [**13**].

We begin our discussion of computational geometry with several basic concepts.

Voronoi Diagrams

Consider the following problem, known as the *post office problem*: We are given a set S of n points in the plane (considered as *post offices* or *sites*). When an arbitrary new point (x, y) (say, a residence) is given, we must find out which post office is closest to (x, y).

In the example above, one expects many queries, with different values of (x, y), to be asked with respect to the same set of post offices. Thus it pays to "preprocess" the set S, that is, to have some useful computation about S done in advance, so that when a query arrives, the answer can be determined quickly. In this particular case, precomputing a map of "postal regions" would seem to be a good idea. For each post office p, the locus of points (x, y) that are closer to p than to any other post office in S is a convex region $V(p)$, called the *Voronoi region* or *Voronoi polygon* associated with p. The n polygons $V(p)$ form a partition of the plane, called the *Voronoi diagram* of S, denoted by $Vor(S)$. (FIGURE 1(b).)



(*a*) A point set S (*b*) $Vor(S)$ (*c*) $DT(S)$

FIG. 1

The earliest significant use of Voronoi regions seems to have occurred in the work of Gauss, Dirichlet and Voronoi [**29**] in their investigations on the reducibility of positive definite quadratic forms (see [**21**], [**17**] for a summary). However, the same concept has evolved independently in a variety of other scientific disciplines. For example, in crystallography, one simple model of crystal growth starts with a fixed collection of sites in 2- or 3-space, and allows crystals to begin growing from each site, spreading out at a uniform rate in all directions, until all space is filled (see [**14**]). The "crystals" (called "Wirkungsbereiche" or "domains of action") then consist of all points nearest to a particular site, and consequently are just the Voronoi regions for the original set of points (FIGURE 2).

Similarly, the statistical analysis of meteorological data led to the formulation (around 1910) of Voronoi regions under the name "Thiessen polygons." The reader is referred to [**17**] for a survey of these applications, and to [**2**] for a more complete discussion of the uses of Voronoi regions in statistics, geography, interpolation, physical chemistry, and many other fields.

The Voronoi diagram enjoys many interesting properties. First of all, the straight-line dual of the Voronoi diagram is a triangulation of S, called the *Delaunay triangulation* and denoted by $DT(S)$ (FIGURE 1(c).) It is easy to see that:

P1. If p is a nearest neighbor of q, then polygons $V(p)$ and $V(q)$ are adjacent in the Voronoi diagram. This means that p and q are connected by an edge in $DT(S)$.

P2. Points p, q, r form a triangle in $DT(S)$ if and only if the circle through p, q, and r has no point of S in its interior.

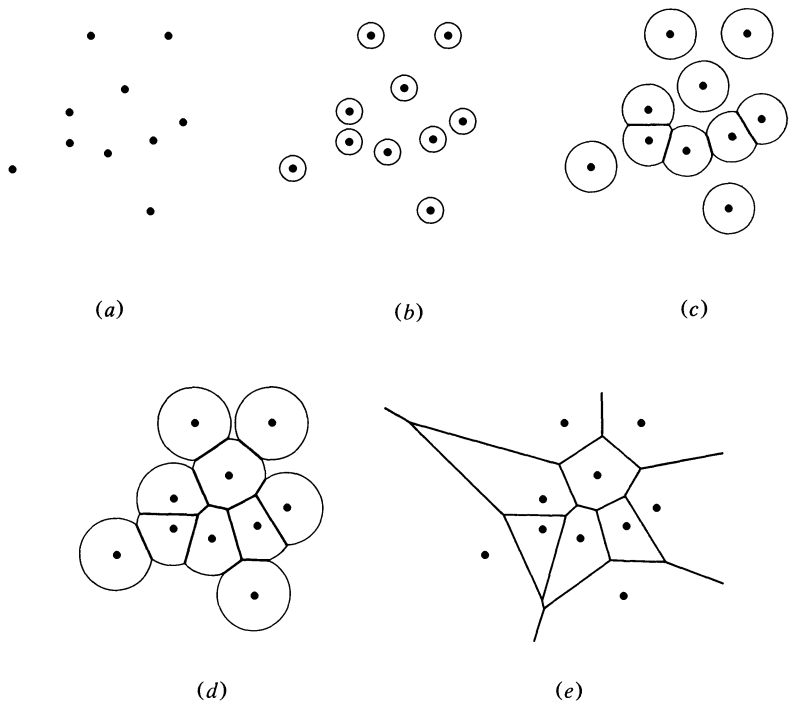


FIG. 2. Model of 2-dimensional crystal growth.

It follows from property P1 that, to find a *closest pair* of points occurring in S, we need only consider the $O(n)$ pairs which are joined by edges in the Delaunay triangulation for S.

Some further useful properties of these diagrams can be established without much difficulty. For example, the Delaunay triangulation contains all *minimum spanning trees* of the point set. A minimum spanning tree (MST) is a tree (i.e., a connected graph with no cycles) that connects all points of V, such that the sum of the edge lengths in the tree is as small as possible. The MST of a point set S may not be unique, but any such tree must use only edges present in $DT(S)$. We state this as:

P3. $MST(S) \subseteq DT(S)$.

Proof. Let $e = (p, q) \in MST(S)$. If $e \notin DT(S)$, then by property P2, the circle with (p, q) as diameter must contain some point of S, say r, in its interior. When we remove the edge e from $MST(S)$, the set S breaks up into two connected components. But the two components can be rejoined by using one of the edges (p, r) or (q, r); since either edge is shorter than the original edge (p, q), we have a contradiction. ■

Property P3 implies that we can find a MST for a set of points by first constructing its Delaunay triangulation $DT(S)$ and then, since the triangulation is a planar graph, apply a particularly efficient (e.g., linear-time) MST algorithm for planar graphs to $DT(S)$.

For a set S of n points in the plane, there may be exponentially many different triangulations. Depending on the application, one may be interested in triangula-

tions that have certain special characteristics. Not surprisingly, the Delaunay triangulation $DT(S)$ satisfies a number of extremal properties, one of which is that it maximizes the minimum angle among all triangulations of S. In other words, $DT(S)$ does the best possible job in avoiding triangles with small angles (see **[11]** for a proof).

Construction of Voronoi Diagrams

How difficult is it to construct the Voronoi diagram $Vor(S)$ from S? A straightforward approach is to compute the polygons $V(p)$ one by one. Write $S = \{p_1, \ldots, p_n\}$. To find $V(p_1)$, say, we can compute, for successively larger values of j, the polygon U_j which is the locus of all points that are closer to p_1 than to any point in $\{p_2, \ldots, p_j\}$. Once U_j is known, we can obtain U_{j+1} by intersecting the convex polygon U_j with the perpendicular bisector between p_1 and p_{j+1}. This step can be accomplished by a fast binary (or logarithmic) search, because the vertices of a convex polygon admit a nice ordering (first increasing and then decreasing in their x-coordinates), and a line can intersect this polygon in at most two places. Thus, the polygon $V(p_1) = U_n$ can be computed in $O(n \log n)$ steps, and therefore the whole diagram $Vor(S)$ in a total of $O(n^2 \log n)$ steps.

To compute $Vor(S)$ more efficiently, we must try to construct several polygons $V(p_i)$ simultaneously. Here we can invoke a general principle, called the *sweep technique*, which is often useful in the design of geometric algorithms. Imagine that a horizontal line is being swept up the plane from $y = -\infty$ to $y = +\infty$. Any cross-section of the two-dimensional diagram $Vor(S)$ as defined by the sweep-line is a simple one-dimensional configuration. Furthermore, the configuration undergoes essential (i.e., topological) changes rather infrequently—only when the sweep-line first enters some new region $V(p_i)$, or has just left such a region; in either case, the change is local. The difficult part, of course, is in knowing when a change should occur, since the sweep-line will enter a region $V(p_i)$ *before* it actually encounters the point p_i. It turns out that this difficulty can be overcome by a clever idea due to Fortune **[12]**. One first performs a non-linear transformation T of the plane, with T depending on the set S. Under this transformation, the new diagram $T(Vor(S))$ has the property that the lowest point of the Voronoi region $T(V(p_i))$ is always at the site $T(p_i)$ itself (FIGURE 3). Thus, we will create a Voronoi region only when the sweep-line reaches a new site, and delete that region when its topmost point is reached. Finally, the real Voronoi diagram can be easily reconstructed from its transformed image. The total running time of this algorithm is $O(n \log n)$, since each insertion/deletion operation can be done in $O(\log n)$ steps with a standard one-dimensional search structure.

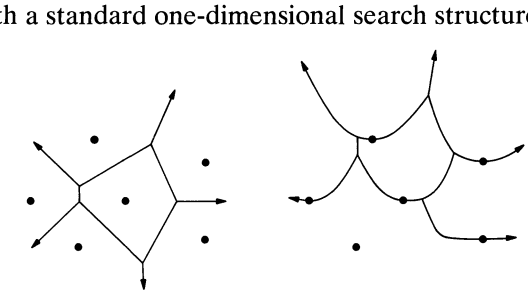


FIG. 3. Fortune's transformation.

The transformation used by Fortune maps a point r in the region $V(p)$ to $r + (0, y)$ where $y = d(r, p)$, the distance from r to p; thus, lines are transformed into hyperbolic arcs. This transformation can be intuitively understood as follows. Think of the ordinary Voronoi diagram as being generated by a physical process in which an *obstacle* is placed at each site, and an expandable disk is moved around so that at any instant, the disk has the maximum possible radius permitted by the obstacles. The locus of the center of the disk, then, defines the Voronoi diagram. If, instead of tracing the center, we trace out the locus of the *topmost* point of the disk at all times, the result would be precisely the transformed Voronoi diagram defined above.

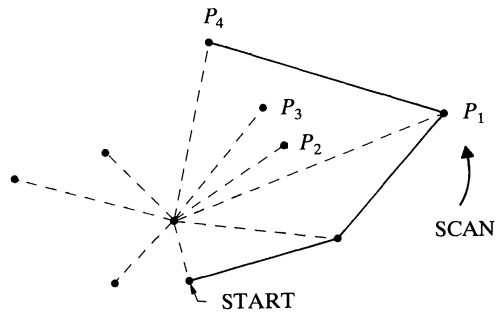
Convex Hulls

It is known that the Voronoi diagram for a finite *convex* planar set, that is, a set of n points forming the vertices of a convex polygon, can be computed in only $O(n)$ time [**1**]. Indeed, convexity is a strong geometric condition that often leads to particularly efficient solutions to computational problems. Consequently, a common strategy used in practice for solving problems on a general point set may involve either first decomposing the set into convex pieces, or approximating the set with its *convex hull*, i.e., the smallest convex set containing the given set. This makes the construction of the convex hull of a set a basic problem in computational geometry.

The first $O(n \log n)$ convex hull algorithm in two dimensions is due to Graham [**15**]. The algorithm first sorts the points by polar angle about an interior point, and then scans the sorted list to discard any point that would cause a reflex angle to occur (FIGURE 4(a)). Note that one may occasionally have to back up a long way and discard many points in order to restore convexity. However, as can be easily argued, the overall cost of the scanning step is still only linear in the total number of points. The convex hull can also be constructed by a divide-and-conquer algorithm ([**25**]). In this approach, we partition the points into two equal subsets P_L and P_R by a vertical line, find their convex hulls recursively, and merge them to form the desired convex hull. The merging step is accomplished by finding the *upper bridge* b_1 and the *lower bridge* b_2 between P_L and P_R (FIGURE 4(b)). To find the upper bridge, we take the line segment l connecting the rightmost point p of P_L and the leftmost point q of P_R, and move its endpoints upward iteratively along the boundaries of P_L and P_R, until l reaches the highest possible position, which now defines b_1. The lower bridge b_2 is found analogously. The time $T(n)$ required satisfies $T(n) \leq 2T(n/2) + O(n)$ which implies $T(n) = O(n \log n)$. If the input points are sorted, or more generally, form the (ordered) vertices of a simple polygon (i.e., a polygon without self-intersections), the convex hull can in fact be found in *linear* time (Graham and Yao [**16**], Lee [**20**]).

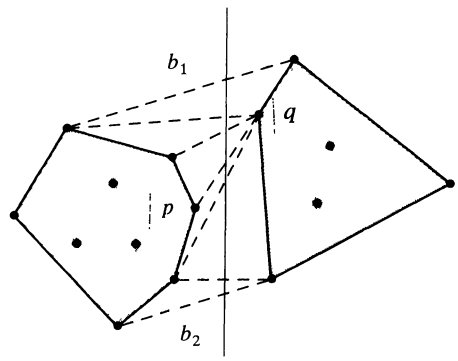
How fast can we compute the convex hull of a set in *three* dimensions? The divide-and-conquer algorithm described above, as it turns out, can be made to work in three dimensions within the same time bound $O(n \log n)$. In the crucial merging step of the algorithm, the hull of the union of two non-intersecting polyhedra can be formed by "gift wrapping" the two objects in one package, so to speak. The merging cost is proportional to the total size of the incidence graphs of the two polyhedra (i.e., the graphs induced by the vertices and edges of the polyhedra), and is $O(n)$ since these incidence graphs are planar [**24**].

For dimensions $d \geq 4$, the situation becomes more complex since a d-polytope with n vertices can have as many as $Cn^{\lfloor d/2 \rfloor}$ faces ([**17**]). There are two types of



Vertex p_2 is eliminated because $\angle p_1 p_2 p_3$ is reflex; then p_3 is eliminated because $\angle p_1 p_3 p_4$ is reflex.

(*a*)
Graham's scan



(*b*)
Divide and conquer

FIG. 4

convex hull algorithms, depending on whether the algorithm only enumerates the *facets*, i.e., the $(d-1)$-faces, of the convex hull or produces the complete *facial lattice*, i.e., a description of all faces and incidence relationships of the convex polytope. Studies of these convex hull algorithms can be found in Chand and Kapur [**6**] and Seidel [**26**]. We note that, through duality, the problem of enumerating the facets of the convex hull of a point set is equivalent to the problem of enumerating the vertices of a polytope defined by a set of linear inequalities; in the latter form this problem has also been widely explored in the optimization literature (see Dyer [**10**] for a survey).

Connection between Voronoi Diagrams and Convex Hulls

Brown [5] was the first to observe that, through an appropriate transform the Voronoi diagram of a set S in R^d corresponds to the convex hull of the transformed set in R^{d+1}. We illustrate this for the simple case $d = 2$. Consider the mapping

$$\alpha: (x, y) \mapsto (x, y, x^2 + y^2)$$

which "lifts" a point (x, y) in the plane onto the paraboloid $z = x^2 + y^2$. Note that cocircular points in the plane are mapped by α into coplanar points in space (FIGURE 5(a)). Given n sites S in the plane, their image points under α form the vertices of a convex polyhedron P (since the paraboloid is convex). This means, for any "downward-looking" face f of the polyhedron P, its defining plane must lie "below" all other vertices of P. Therefore, by our earlier observation, the projection of f corresponds to a circle with no points of S in its interior. This implies that the projection of all the downward-looking faces of the polyhedron P gives precisely the faces of the Delaunay triangulation of the planar set S (FIGURE 5(b)). This relationship readily generalizes to higher dimensions, and enables one to use convex hull algorithms in $d + 1$ dimensions to obtain Voronoi diagrams in d dimensions!

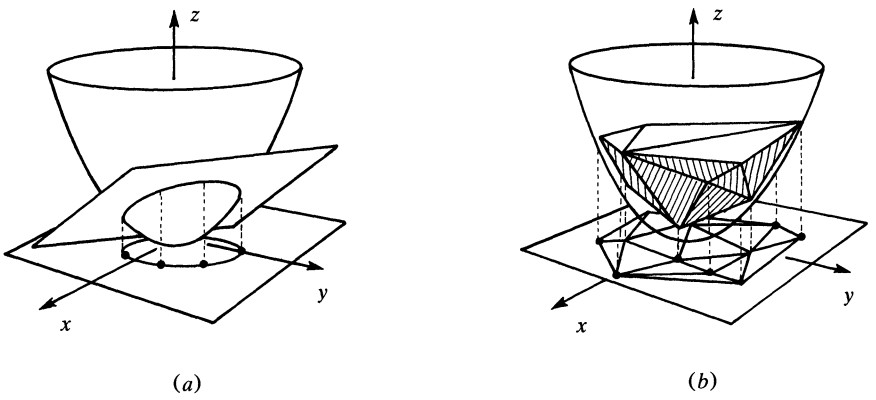


FIG. 5. Voronoi diagrams and convex hulls.

Combinatorial Geometry

The analysis of geometric algorithms often requires detailed combinatorial knowledge of the geometric structures involved. The book by Edelsbrunner [11] discusses many topics which overlap both computational and combinatorial geometry. We give one such example here.

Many algorithms in computational geometry involve calculation of the lower envelope (i.e., pointwise minimum) of a collection of functions. For example, suppose we are to render the 2-dimensional image of a 3-dimensional *terrain*, which is a polyhedral surface with the property that any line parallel to the z-axis intersects the surface at most once. From any viewpoint v, the "upper rim" (also called the *silhouette*) of the terrain as seen from v can be expressed as the lower envelope of a set of line segments in two dimensions (FIGURE 6).

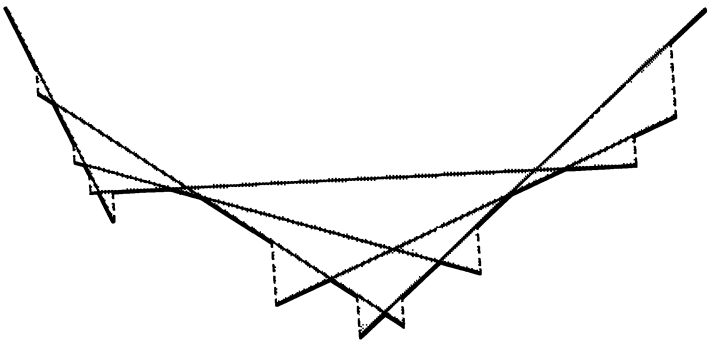


FIG. 6. The lower envelope of a set of line segments.

Lower envelopes are closely related to *Davenport-Schinzel sequences* (or *DS* sequences, for short). Specifically, an (n, s)-Davenport-Schinzel sequence is a sequence composed of n symbols satisfying: (i) no two adjacent elements are equal, and (ii) the sequence does not contain as a (not necessarily consecutive) subsequence of length $s + 2$ any alternation of two distinct symbols. The parameter s is called the *order* of the *DS* sequence. In the above example, the lower envelope of a set of n segments $\{l_1, \ldots, l_n\}$ in the plane is a *DS* sequence of order 3, since no alternation of the form $\cdots l_i \cdots l_j \cdots l_i \cdots l_j \cdots l_i \cdots$ can occur in the lower envelope. It has recently been shown that the maximum length $\lambda_3(n)$ of an $(n, 3)$-*DS* sequence is $\lambda_3(n) = \Theta(n\alpha(n))$, where $\alpha(n)$ is the extremely slowly growing functional inverse of the Ackermann function ([**19**]). In fact, this maximum value can be realized by a rather intricate arrangement of n line segments [**30**]. Thus the silhouette of a general polyhedral surface has complexity $\Theta(n\alpha(n))$, and can be computed in $O(n\alpha(n)\log n)$ time [**8**]. Other geometric problems where Davenport-Schinzel sequences naturally arise include the calculation of Euclidean shortest paths in 3-space amidst polyhedral obstacles, hidden-surface removal from a varying viewpoint, and the computation of other time-varying geometric configurations ([**4**], [**27**]).

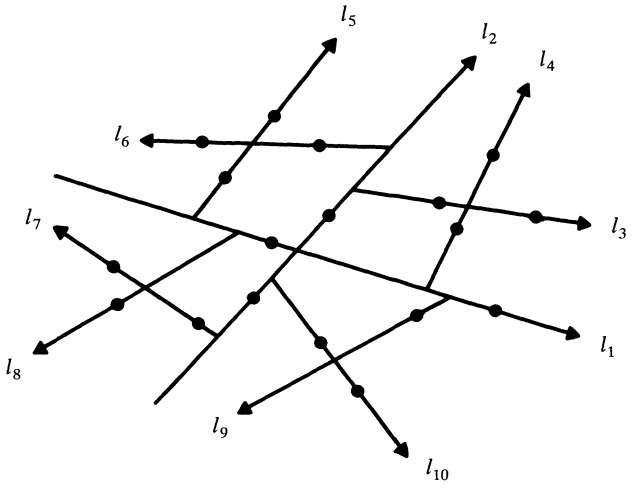
Space Partitions and Range Search

The post office problem mentioned earlier belongs to a class of problems studied in computational geometry known as "range search problems". In the most general setting, the database S to be searched may contain higher order objects than points, and the queries can be of a more complex nature. Still, let us confine our discussion here to the case when S is a set of points in R^d. An important type of query is the "half-space" query, in which an arbitrary linear inequality is specified, and all the points in S that satisfy the inequality should be identified quickly. As an example of a half-space query, consider a database where the first two coordinates contain, respectively, the income and the number of children of a household, and a query asks for all households with monthly income exceeding \$1000 plus \$200 for each child.

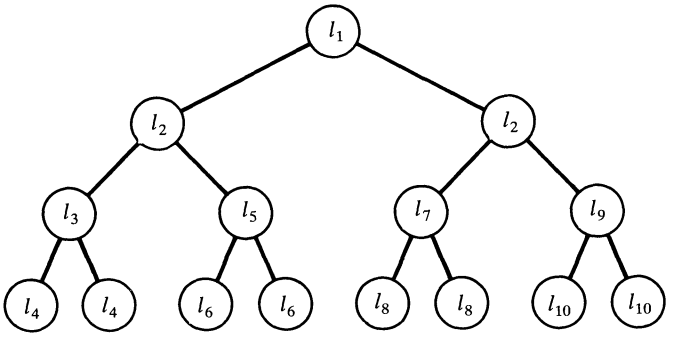
The first nontrivial solution to this problem was proposed by Willard [**31**]. It makes use of the following well-known fact: a density function over a bounded region can be partitioned by two straight lines l_1, l_2 into four quadrants with equal

mass. Translating this result into the discrete case, it says that a set S of n points in the plane can be partitioned by two lines into four (open) quadrants such that at most $n/4$ points are contained in the interior of each quadrant. One can then further partition the points in each quadrant, and repeat this process until all points are separated. This gives a tree structure for representing the set S (see FIGURE 7). For a query half-plane Q, since the line defining Q can intersect at most three of the four quadrants in each partition, one of the quadrants is either entirely inside or outside of Q, and hence can be exempted from further search. The search time $T(n)$ thus satisfies a recurrence relation $T(n) = 3T(n/4) + O(1)$, which yields $T(n) = O(n^{\alpha})$ with $\alpha = \log_4 3 \cong 0.774$.

Can we extend such 4-*partitions* in R^2 to 8-*partitions* in R^3? The answer turns out to be yes. Given a set S of n points in R^3, one can always find three planes



(*a*)
A 4-partition for 14 points



(*b*)
The corresponding tree structure

FIG. 7

that form an 8-partition in the sense that at most $n/8$ points of S lie in each of the eight open regions. Thus, we can represent a point set in R^3 by recursive 8-partitions. Since an arbitrary plane in R^3 can intersect at most seven of the eight open regions defined by an 8-partition, this leads to a retrieval time of $O(n^\alpha)$ with $\alpha = \log_8 7 \cong 0.936$ for half-space queries.

The 8-*partition* proposition stated above is considerably more difficult to prove than the two-dimensional analogue [**34**]. However, we note that a weaker version of it actually suffices to guarantee the same query time $O(n^{\log_8 7})$. In this version, the partition consists of a bisecting plane H for S, together with two *separate* 4-partitions for the sets S^+, S^- above and below H, with the condition that the half-axes of the two partitions are lined up exactly (FIGURE 8).

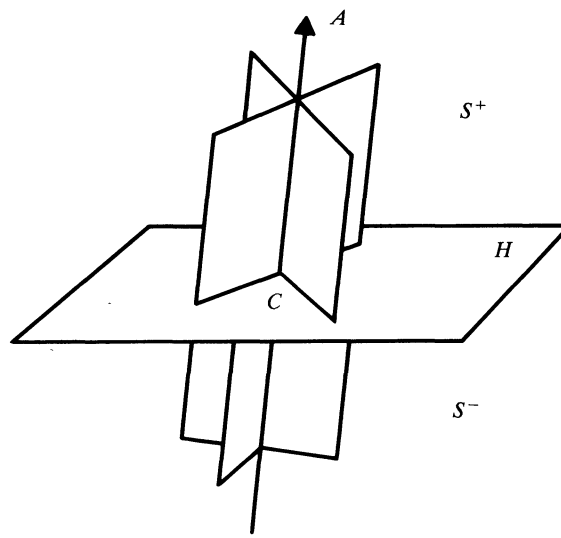


FIG. 8. An 8-partition with axis A and center C.

To show that such a partition exists, we invoke the well known Borsuk-Ulam theorem.

THEOREM (Borsuk-Ulam, [**23**]). *Let f: $S^d \mapsto R^d$ be a continuous, antipodal map, i.e., $f(-\nu) = -f(\nu)$ for $\nu \in S^d$, the unit d-sphere. Then there is a point $\nu \in S^d$ such that $f(\nu) = 0$.*

To obtain the desired partition, consider $p_\nu(S^+)$, the projection of S^+ onto H along direction ν, where ν is any vector in the southern hemisphere of S^2. We find a (unique) 4-partition (L_1, L_2) for the planar set $p_\nu(S^+)$ with the constraint that L_1 is parallel to the x-axis. Call the intersection point of L_1 and L_2 the *center* of $p_\nu(S^+)$, denoted by $C(\nu)$. Similarly define the center $C'(\nu)$ for $p_{-\nu}(S^-)$, which is the projection of S^- along the direction $-\nu$. Applying the Borsuk-Ulam theorem for $d = 2$, it follows that the function $f(\nu) = C(\nu) - C'(\nu)$ must vanish for some vector ν_0. This gives us what we want: separate 4-partitions for S^+ and S^- which share a common axis A, where A has direction ν_0 and goes through the point $C(\nu_0) = C'(\nu_0)$. See [**32**] for complete details.

It is easy to see that the resulting 8-partition does have the property that any query plane can intersect at most 7 out of the 8 regions. Furthermore, this version can be generalized to higher dimensions by an inductive argument, thus allowing us to solve half-space queries in any number of dimensions.

It is an interesting open question whether the (strong) 8-partition proposition can be generalized to four dimensions; that is, given a density function over a bounded region in four dimensions, whether it can always be partitioned by 4 hyperplanes into 16 orthants with equal mass. However, it is not hard to show that (strong) 2^d-partitions are not always possible in dimensions $d \geq 5$. For example, take thirteen small balls of equal mass in R^5, and let the balls be in general position, so that no hyperplane can intersect more than five of them. Then, in any 2^5-partition, each ball must be cut by at least two hyperplanes, since otherwise some orthant would contain at least half of a ball, and thus at least $1/26$ of the total mass, which is more than the required fraction $1/32$. Therefore, for all thirteen balls, at least 26 instances of hyperplane-ball intersections are needed. Since the balls are in general position, five hyperplanes can provide at most 25 such instances and we have a contradiction.

Lower Bounds

An important, and often difficult, challenge in algorithm design is to know whether an algorithm can still be improved upon, or in fact one has already found the fastest algorithm possible. To resolve such questions, we must establish *lower bounds* to the computational complexity of a problem in some fairly general model.

For geometric problems, a natural model for studying computational complexity is the *algebraic computation tree model*. In this model, an algorithm can in one step either perform an arithmetic operation, or make a comparison and then branch according to the result of the comparison. (Thus, each comparison in fact corrresponds to a polynomial test of arbitrary degree evaluated at the input.) FIGURE 9 gives an example of an algebraic computation tree for testing whether a point (x, y) lies in the region

$$W = \{(x, y) | xy > 1, -2.5 < x + y < 3\}.$$

Let us consider the problem of finding a closest pair among n points in the plane. As we saw before, this problem can be solved in $O(n \log n)$ time by first constructing the Voronoi diagram and then performing a search in linear time. It turns out that $\Omega(n \log n)$ is also a lower bound for this problem in the algebraic computation tree model. In fact, the simpler problem of *deciding* whether the input set contains a *repeated* point, i.e., whether $p_i = p_j$ for some $i \neq j$, has complexity $\Omega(n \log n)$ even in *one* dimension! This result is due to Ben-Or [3], and the main idea of the proof is the following. The upper bound follows simply by sorting the n points and testing adjacent pairs for equality. The lower bound is obtained by mapping the polynomial inequalities to a suitable algebraic variety and then associating the number of computational steps with the Betti numbers of the algebraic variety. Let $x = (x_1, x_2, \ldots, x_n) \in R^n$ denote the input, and let T be an algebraic computation tree that decides whether $x \in W$ where W is defined by

$$W = \left\{ (x_1, x_2, \ldots, x_n) \,\middle|\, \prod_{i \neq j} (x_i - x_j) \neq 0 \right\} \subseteq R^n.$$

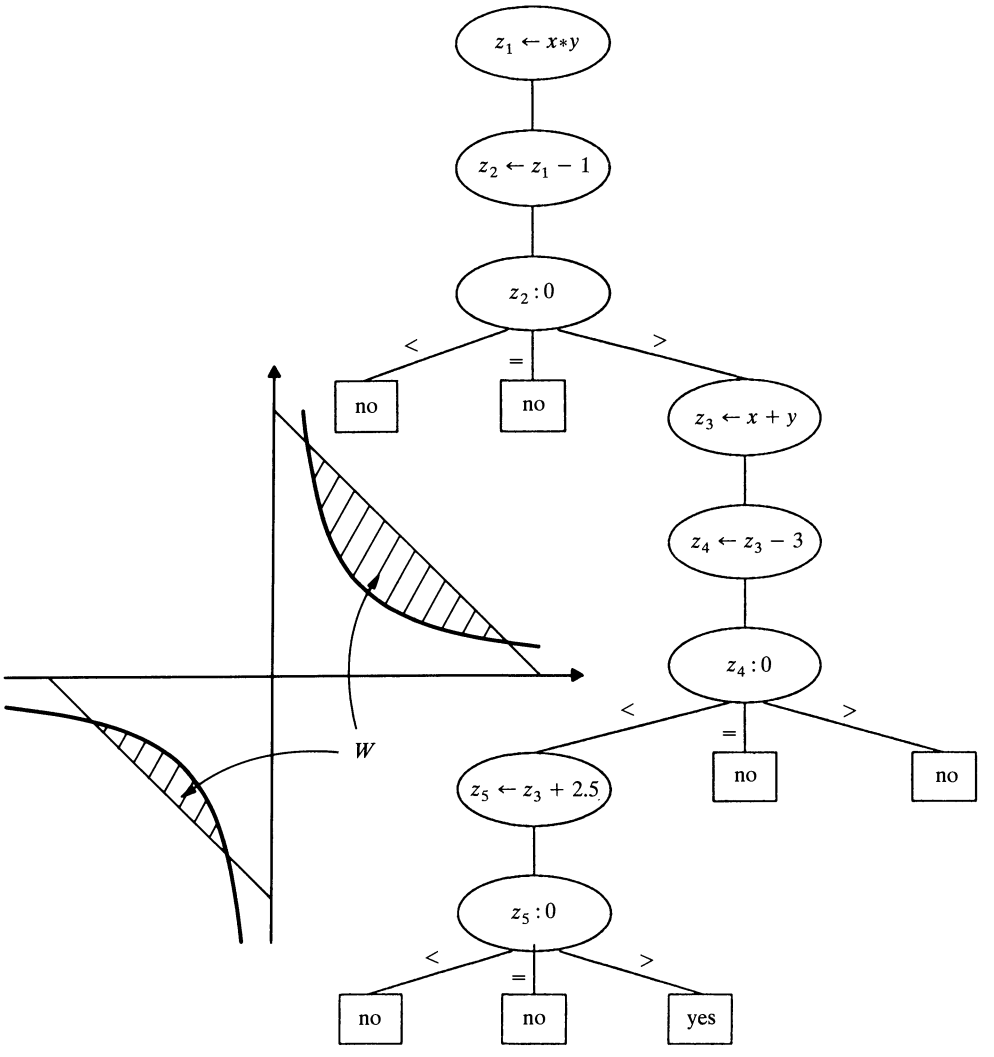


FIG. 9. An algebraic computation tree for testing "Is $(x, y) \in W$."

It is easy to see that W has $n!$ connected components. At the same time, by invoking a theorem of Milnor-Thom ([22], [28]), one can show that the algebraic variety V associated with a leaf of T has at most 3^{n+h} components, if the path to that leaf involves h operations. For a leaf associated with a "yes" answer, each connected component of V must be completely contained in some connected component of W. We thus conclude that the maximum path length h has to satisfy $3^h 3^{n+h} \geq n!$, which implies $h = \Omega(n \log n)$.

Common Techniques

As we have seen, computational geometry uses concepts and results from classical geometry, topology, combinatorics, as well as standard algorithmic techniques such as sorting and searching, graph manipulations, and linear program-

ming. In addition, certain special techniques and paradigms have emerged and found repeated application in computational geometry. We briefly summarize some of the most common ones below. Much more comprehensive surveys can be found in [11], [25] and [33].

1. Divide and conquer. The basic idea is to divide the problem into two or more subproblems, solve them recursively and then merge the results to obtain a solution to the original problem. Although this technique is common in algorithm design, it seems particularly well suited for solving geometric problems: the dividing step can naturally be accomplished by splitting along a line (or a hyperplane in d dimensions), and the merging step can take advantage of reduced problem complexity across the dividing hyperplane. The convex hull algorithms in two and three dimensions are good examples.

2. The sweep technique. This is a technique designed to reduce the dimension of a (static) geometric problem at the cost of changing its mode from static to dynamic. For example, in the construction of Voronoi diagrams, the sweep-line induces one-dimensional cross-sections which change dynamically (but only locally) when the sweep-line moves from one (discrete) position to the next. The sweep technique is also quite useful in higher dimensions.

3. Geometric transformations. Through a geometric transformation, one can often state a given problem in an alternative form which, though equivalent, may shed new light on the problem. One of the transforms used most often in computational geometry is the *dual transform* between points and hyperplanes. Other examples include the hyperbolic transform of Fortune for Voronoi diagrams and the correspondence between Voronoi diagrams and convex hulls (both discussed earlier), the use of *Plücker* coordinates to represent a line in three dimensions as a point in five dimensions, and the *skewed projection* determined by a pair of lines in three dimensions.

4. The locus approach. This approach applies to geometric query problems and it works as follows. We interpret the query as a point in some space (through an initial transformation, if necessary), and then partition that space into a set of non-overlapping regions so that the answer is invariant for all query points that fall in the same region. Several ingredients are involved in this approach: transformation of the query into a point, definition and construction of a suitable space partition, and an algorithm for point-location. Using the Voronoi diagram to solve the nearest-neighbor query is a basic example of this approach.

5. Random sampling. The idea is to use random sampling of the input objects to carry out divide-and-conquer efficiently, that is, to split up the original problem into subproblems each guaranteed to be of small size [7]. The expected performance of the resulting algorithm is with respect to the built-in randomization process, and does not depend on any assumption about the input distribution. This technique often yields simple algorithms with good expected performance.

Conclusion

We have tried here to give a brief sketch of some of the main ideas underlying the dynamically growing field of computational geometry. It is the natural convergence of ideas from many areas of mathematics such as topology, combinatorics,

algebra, probability and, of course, geometry, with those from computer science, such as graph algorithms, data structures, and optimization. We feel confident that the current trend of studying computational questions on geometric objects will continue to suggest new classes of problems which in turn will enrich and become a permanent part of the glorious tradition of geometry.

REFERENCES

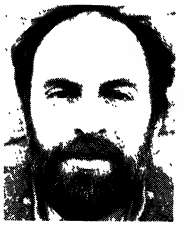
1. A. Aggarwal, L. J. Guibas, J. Saxe and P. Shor, A linear time algorithm for computing the Voronoi diagram of a convex polygon, *Proc. 19th Ann. ACM Symp. Theory Comput.*, (1987) 39–45.
2. F. Aurenhammer, Voronoi diagrams—a survey, Inst. for Inf. Prc., Graz Tech. Univ. Tech. report, 1988.
3. M. Ben-Or, Lower bounds for algebraic computation trees, *Proc. 15th Ann. Symp. Theory Comput.*, (1983) 80–86.
4. M. Bern, D. Dobkin, D. Eppstein and R. Grossman, Visibility with a moving point of view, *Proc. 1st Ann. ACM-SIAM Symposium on Discrete Algorithms* (1990) 107–117.
5. K. Q. Brown, Voronoi diagrams from convex hulls, *Inform. Process. Lett.*, 9 (1979) 223–228.
6. D. R. Chand and S. S. Kapur, An algorithm for convex polytopes, *J. ACM*, 17 (1970) 78–86.
7. K. L. Clarkson, New applications of random sampling in computational geometry, *Discrete and Comput. Geometry*, 2 (1987) 195–222.
8. R. Cole and M. Sharir, Visibility problems for polyhedral terrains, *J. Symbolic Computation*, 7 (1989) 11–30.
9. M. Davis, Y. Matijasevič and J. Robinson, Hilbert's Tenth Problem. Diophantine Equations: Positive aspects of a negative solution, in Mathematical developments arising from Hilbert Problems (F. Browder, ed.), *Proc. Symp. Pure Math 28, part 2*, (1976), 323–378.
10. M. E. Dyer, The complexity of vertex enumeration methods, *Math. Oper. Res.*, 8 (1983) 381–402.
11. H. Edelsbrunner, Algorithms in Combinatorial Geometry, Springer-Verlag, Heidelberg, Germany, 1987.
12. S. J. Fortune, A sweepline algorithm for Voronoi diagrams, *Proc. 2nd Ann. ACM Symp. Comput. Geom.* (1986) 313–322.
13. M. R. Garey and D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-completeness, W. H. Freeman and Co., San Francisco, 1979.
14. E. N. Gilbert, Random subdivisions of space into crystals, *Annals of Math. Stat.*, 33 (1962) 958–972.
15. R. L. Graham, An efficient algorithm for determining the convex hull of a finite planar set, *Inform. Process. Lett.*, **1** (1972) 132–133.
16. R. L. Graham and F. F. Yao, Finding the convex hull of a simple polygon, *J. Algorithms*, 4 (1983) 324–331.
17. B. Grünbaum and G. C. Shephard, Tilings and Patterns, W. H. Freeman and Co., New York, 1987.
18. L. J. Guibas and J. Stolfi, Ruler, compass and computer: The design and analysis of geometric algorithms, DEC Sys. Res. Center Tech. Dep. 37 (1989), 55 pp.
19. H. Hart and M. Sharir, Nonlinearity of Davenport-Schinzel sequences and of generalized path compression schemes, *Combinatorica*, 6 (1986) 151–177.
20. D. T. Lee, On finding the convex hull of a simple polygon, *Int. J. Comput. Inform. Sci.*, 12 (1983) 87–98.
21. C. G. Lekkerkerker, Geometry of Numbers, Wiley Interscience, New York, 1969.
22. J. Milnor, On the Betti numbers of real algebraic varieties, *Proc. AMS*, 15 (1964) 275–280.
23. J. R. Munkres, Elements of Algebraic Topology, Addison-Wesley, Reading, MA, 1984.
24. F. P. Preparata and S. J. Hong, Convex hulls of finite sets of points in two and three dimensions, *Comm. ACM*, 2 (1977) 87–93.
25. F. P. Preparata and M. I. Shams, Computational Geometry: An Introduction, Springer-Verlag, New York, 1985.
26. R. Seidel, A convex hull algorithm optimal for point sets in even dimensions, Univ. of B.C., CS Tech Rep. 81–14, 1981.
27. J. T. Schwartz, M. Sharir and J. Hopcroft, Planning, Geometry and Complexity of Robot Motion, Ablex, Norwood, NJ, 1986.

28. R. Thom, Sur l'homologie des variétés algébriques réelles. Differential and Combinatorial Topology (ed., S. S. Cairns), Princeton Univ. Press, 1965.
29. M. G. Voronoi, Nouvelles applications des parametres continus a la théorie des formes quadratiques, *J. Reine u. Agnew. Math.*, 134 (1908) 198–287.
30. A. Wiernik and M. Sharir, Planar realizations of nonlinear Davenport-Schinzel sequences, *Discrete and Comput. Geometry* 3 (1988), 15–48.
31. D. E. Willard, New data structures for orthogonal range queries, *SIAM J. Computing*, 14 (1985) 232–253.
32. A. C. Yao and F. F. Yao, A general appraoch to d-dimensional geometric queries, Proc. 17th Ann. ACM Symp. on Theory of Computing, 1985, 163–168.
33. F. F. Yao, Computational Geometry (to appear as a chapter in Handbook of Theoretical Computer Science, North Holland, Amsterdam, 1990).
34 F. F. Yao, D. P. Dobkin, H. Edelsbrunner and M. S. Paterson, Partitioning space for range queries, *SIAM J. Computing*, 18 (1989) 371–384.

Minimal Surfaces Based on the Catenoid

DAVID HOFFMAN* and WILLIAM H. MEEKS, III*, *University of Massachusetts, Amherst, MA 01003*

DAVID HOFFMAN is Professor of Mathematics and Co-Director of the Geometry, Analysis, Numerics and Graphics Center (GANG) at the University of Massachusetts, Amherst. He earned his Ph.D. in mathematics at Stanford, after receiving undergraduate degrees at the University of Rochester (in history and mathematics). He has pursued research and/or teaching at the Universities of Durham and Warwick (UK), Michigan and Stanford, as well as IMPA (Rio de Janeiro, Brazil) and the University of Paris VII. He is the 1990 recipient of the MAA Chauvenet Prize.



WILLIAM MEEKS received his Ph.D. from Berkeley in 1975, under the direction of Blaine Lawson. He has held positions at IMPA, Rice, UCLA, Stanford, and the Institute for Advanced Study. His research interests lie in differential geometry, minimal surfaces, low-dimensional topology, and computational methods. He has given invited addresses at MSRI, at the 1985 Annual Meeting of the AMS, and at the ICM in Berkeley, 1986.



Contents

1. The flux formula, the catenoid and embedded minimal annuli 705
2. Complete embedded minimal surfaces with finite total curvature 707
 2.1. New finite total curvature examples 711
 2.2. More finite total curvature examples 714
3. Minimal surfaces foliated by convex curves in parallel planes 716
 3.1. Minimal annuli 716
 3.2. Schiffman's First and Second Theorem and extensions 716
 3.3. Minimal surfaces foliated by circles and lines in parallel planes: Riemann's examples 719
4. Embedded periodic minimal surfaces and the period problem 723
5. Conclusion 726
 5.1. The topology of proper periodic minimal embeddings in R^3 726
 5.2. Periodic interfaces in materials as a source of new minimal surfaces 726

In this article we will focus on the subject of embedded minimal surfaces. We will try to illustrate how some of the modern-day work in the subject—both in construction of examples and in development of a global theory—is dependent

*The research described in this paper was supported by research grant DE-FG02-86ER25015 of the Applied Mathematical Science subprogram of the Office of Energy Research, U.S. Department of Energy, and National Science Foundation grant DMS-8802858, DMS-8810088 and DMS-8900285.

upon understanding the catenoid. As we will see in the next section, the catenoid is the unique nonplanar minimal surface of revolution, and so is really the simplest complete minimal surface, after the plane. Nonetheless, we will try to show that it is the most important example.

A surface S in $\mathbb{R}^3$ is said to be *minimal* if each point has a neighborhood $\mathscr{O} \subset S$ that is the surface of least area for its boundary $\partial\mathscr{O}$. Such a surface may have infinite area, the flat plane for example. Alternatively, a minimal surface is one for which every compact subdomain $\mathscr{C}$ is a critical point for area among surfaces with boundary equal to $\partial\mathscr{C}$. From the point of view of local geometry, this is equivalent to the requirement that the mean curvature H is identically zero. The mean curvature is the average of the principal curvatures, so $H = 0$ means that at each point the *average normal curvature* is zero. From a physical perspective, loosely speaking, one imagines the surface as made up of very many rubber bands, stretched in all directions; on a minimal surface the forces due to the rubber bands balance out, and the surface does not need to move to reduce tension. This is not to say that a minimal surface is necessarily *stable*[1]. In fact this is not the case for most of the surfaces. For example, among complete immersed minimal surfaces in $\mathbb{R}^3$, the only stable orientable example is the flat plane [**18, 22**].

One of the most important facts about minimal surfaces is that their coordinate functions are harmonic[2]. Harmonic coordinates actually characterize minimal surfaces, as can be seen from the following general relation

$$\Delta_B X = 2H\vec{N},$$

where $X\colon M \to \mathbb{R}^3$ is an isometric immersion and $\vec{N}$ is the oriented unit-normal vector field associated to the immersion. When $H = 0$, it follows that any linear function on $\mathbb{R}^3$, when restricted to M, is harmonic:

$$\Delta_B X_v = 0,$$

where $X_v(p) = \vec{v} \cdot p$.

In fact, notice that if a surface has a single coordinate function X_v that is harmonic, then combining the previous two formulae in this special case gives

$$0 = \Delta_B X_v = \Delta_B X \cdot v = 2H\vec{N} \cdot v.$$

[1]By stable, we mean that each compact subdomain is a local minimum for area among surfaces sharing the same boundary.

[2]A function $f\colon S \to \mathbb{R}^n$ on a surface is harmonic if

$$\Delta_B f = 0$$

in the induced metric, where Δ_B is the Laplace-Beltrami operator. In conformal coordinates,

$$ds^2 = \lambda\left(du_1^2 + du_2^2\right) \qquad \Delta_B = \lambda^{-1}\Delta,$$

where

$$\Delta = \left(\frac{\partial^2}{\partial u_1^2} + \frac{\partial^2}{\partial u_2^2}\right)$$

is the usual Laplacian.

In a neighborhood of any point where the Gauss normal mapping is nondegenerate, we must have $H = 0$. Thus on the open set where the normal mapping is nondegenerate, H vanishes. Assuming the surface is connected, this means that either the surface is minimal or the Gauss map is of rank less than 2 everywhere, which implies it is flat ($K = 0$). Simple analysis will show that this latter situation can only happen if the surface is a cylinder over a plane curve. Thus, among surfaces with nondegenerate Gauss mappings, minimal surfaces are characterized by the existence of a *single* harmonic coordinate function.

The power of this analytical characterization becomes evident if we work with conformal coordinates (u_1, u_2); that is, coordinates in terms of which the metric looks like $ds^2 = \lambda(du_1^2 + du_2^2)$. Let

$$z = u_1 + iu_2, \qquad \frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial u_1} - i\frac{\partial}{\partial u_2}\right) \quad \text{and} \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial u_1} + i\frac{\partial}{\partial u_2}\right).$$

Then for any $f\colon M \to \mathbb{R}^n$

$$\lambda \Delta_B f = \Delta f = 4\frac{\partial f}{\partial z \partial \bar{z}}.$$

For a minimal surface, this means that

$$\frac{\partial}{\partial \bar{z}}\left(\frac{\partial X}{\partial z}\right) = 0;$$

i.e., the locally defined map $\dfrac{\partial X}{\partial z}$ from M to $\mathbb{C}^3$ is holomorphic.[3] But

$$\frac{\partial X}{\partial z} = \frac{1}{2}\left(\frac{\partial X}{\partial u_1} - i\frac{\partial X}{\partial u_2}\right),$$

a complex 3-vector whose real and imaginary parts consist of real 3-vectors forming an orthogonal basis for the tangent space of $X(M)$ at each point. Since

$$\vec{N} = \lambda^{-1}\left(\frac{\partial X}{\partial u_1} \times \frac{\partial X}{\partial u_2}\right),$$

we may use this information to prove that the Gauss mapping of X is a meromorphic mapping to S^2. Let $\sigma\colon S^2 \to \mathbb{C} \cup \{\infty\}$ be stereographic projection. If we let $g = \sigma \circ N$ be the stereographic projection of the Gauss map, it follows that g is conformal if and only if X is minimal.

We have seen that on a minimal surface M every linear function on $\mathbb{R}^3$ restricts to a harmonic function on M. Their conjugate harmonic functions are not, in general, globally defined. But the differentials of the conjugate harmonic functions are, which means that every linear function gives rise to a holomorphic one-form

[3] We may consider M as a Riemann surface, with complex coordinates given by local conformal coordinates.

on M. One of the most important tools in the classical study of minimal surfaces, the Enneper-Weierstrass representation, makes essential use of this analytic data. (See FIGURE EW.) In fact, given a meromorphic function and a holomorphic one-form on a Riemann surface they determine, *locally*, a minimal surface. The *global* use of the Enneper-Weierstrass representation will be central in our later discussion.

1. The flux formula, the catenoid and embedded minimal annuli

A simple but very useful consequence of this analytical characterization of minimal surfaces by harmonic coordinate functions is the flux formula. According to Stoke's Theorem, for any C^2-function f: $S \to \mathbb{R}^n$

$$\int_S \Delta_B f \, dA = \int_{\partial S} df(\vec{n}) \, ds, \tag{1.1}$$

where dA is the element of area on S, ds is arc-length on ∂S, $\vec{n}$ is the outward pointing conormal to S along ∂S and $df(\vec{n})$ is the directional derivative of f in the direction $\vec{n}$. Applying (1.1) to an isometric immersion X: $S \to \mathbb{R}^3$, we have that $dX(\vec{n})$ is the image in $\mathbb{R}^3$ of the outward-pointing conormal; writing $n^* = dX(\vec{n})$ we have $\int_s \Delta_B f \, dA = \int_{\partial s} n^* \, ds$. (See FIGURE 1.) If X is minimal we have

$$\int_{\partial S} n^* \, ds = 0. \tag{1.2}$$

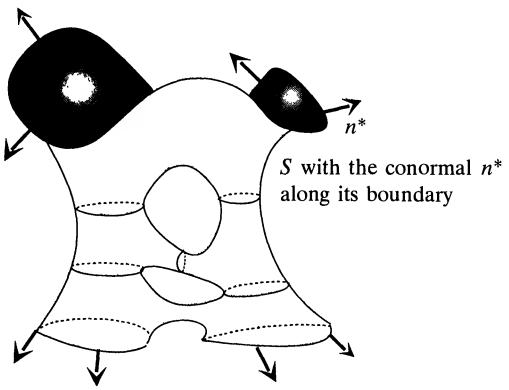


FIG. 1

In particular, if $\vec{v}$ is any fixed vector in $\mathbb{R}^3$

$$\int_{\partial S} n^* \cdot \vec{v} \, ds = 0. \tag{1.3}$$

The integral in (1.3) can be thought of as the flux through ∂S of the flow in $\mathbb{R}^3$ with constant velocity vector $\vec{v}$. While (1.2) and (1.3) are quite simple and were undoubtedly known in the 19th century, they have only recently come into

widespread use in the study of minimal and constant mean curvature surfaces [39, 40].

As a sample application of the flux formula, we consider the catenoid. It was Euler who discovered the catenoid, which was the first nonplanar example of a minimal surface. He did this by finding the surface of revolution that was a critical point for area. Consider a surface of revolution about the z-axis with profile curve $(r(t), t)$ in the (x, z)-plane. Let S be the compact portion S of the surface that is between $z = t_1$ and $z = t_2$. S is bounded by two circles of radii $r(t_1)$ and $r(t_2)$, respectively. Let $\alpha(t)$ be the angle between the tangent plane to S at level $z = t$ and the plane $\{z = t\}$. Then computing the flux in the z-direction ($\vec{v} = (0, 0, 1)$), gives by (1.3)

$$\int_{S\cap\{z=t_1\}} \sin\alpha(t_1)\, ds = \int_{S\cap\{z=t_2\}} \sin\alpha(t_2)\, ds$$

or $2\pi r(t_1)\sin\alpha(t_1) = 2\pi r(t_2)\sin\alpha(t_2)$. But t_1 and t_2 are arbitrary, so $r(t)\sin\alpha(t) = c$, or

$$r(t)\frac{1}{\sqrt{1 + r'^2(t)}} = c. \tag{1.4}$$

This ordinary differential equation is satisfied by the functions

$$r(t) = \frac{\cosh(At + B)}{A}, \qquad A \neq 0, \tag{1.5}$$

and these are all the solutions. Thus, nonplanar minimal surfaces of revolution are all *catenoids*[4]. Note that two catenoids corresponding to different solutions of (1.5) differ by a homothety and translation, and so there is really only one nonplanar minimal surface of revolution. The catenoid has the following properties:

C1) It meets each plane $\{z = t\}$ in a circle;

C2) Topologically it is an annulus; conformally it is a sphere minus two points;

C3) Its Gauss map N is one-to-one and onto the sphere minus the poles;

C4) The portion of the catenoid outside of a sufficiently large compact set of $\mathbb{R}^3$ consists of two unbounded components.

When one tries to look at minimal surfaces similar to but more general than the catenoid, things get very much harder. For example, J. C. C. Nitsche [49] has proved that any minimal annulus that meets *every* $\{z = t\}$ plane in a single star-shaped curve must be the catenoid. He asked whether "star-shaped curve" could be weakened to "Jordan curve."

Q1) Is there a minimal annulus other than the catenoid that meets every horizontal plane in a single Jordan curve?

The affirmative answer to this question is known as the "Nitsche Conjecture." However, it is still an open problem. More generally, the catenoid is the only

[4]The name derives from the fact that these profile curves are also the solutions to the differential equation for the shape of a hanging chain, a catenary, namely (1.4).

Enneper-Weierstrass Representation

Given a Riemann surface M, a meromorphic function g and a holomorphic one-form η on M, there exists a conformal minimal mapping

$$X = (X_1, X_2, X_3)$$

whose Gauss map N has stereographic projection g and whose third coordinate function satisfies

$$\frac{\partial X_3}{\partial \zeta}\, d\zeta = 2g\eta.$$

Specifically

$$X(p) = \operatorname{Re}\int_{p_0}^{p} \Phi, \qquad \text{where}$$

$$\Phi = \left(1 - g^2, i\left(1 + g^2\right), 2g\right)\eta.$$

The metric induced by this mapping is given by

$$\lambda^2 |d\zeta|^2, \qquad \text{where}$$

$$\lambda = 2|\eta|\left(1 + |g|^2\right).$$

Thus X is regular if and only if the zeros of η (order $2m$) occur precisely at the poles (order m) of g. Isolated branch points may occur where this condition is violated.

The metric is complete provided

$$\int_{\delta} \lambda\, ds = \infty$$

for any divergent curve δ on M.

For any nontrivial closed curve $\gamma \subset M$, there is an associated period vector

$$\vec{T}_\gamma = \operatorname{Re}\int_{\gamma} \Phi.$$

If any $T_\gamma \neq 0$, then the mapping X is not well defined on M, but only on some covering of M. To produce a well-defined immersion of M, it is required that

$$\vec{T}_\gamma = 0$$

for all closed cycles on M.

Perodic surfaces may be constructed by having one or more nonvanishing periods.

FIG. EW

known complete, properly[5] embedded minimal surface that is topologically an annulus.

One would like to know all the properly embedded minimal surfaces of finite topology. A surface is said to have *finite topology* if it is homeomorphic to a compact surface from which a finite number of points have been removed. If such a surface is properly embedded in $\mathbb{R}^3$, then each of these points corresponds to a distinct way to diverge in $\mathbb{R}^3$. The image of a small disk neighborhood of a puncture point is called an *end* of the surface. The catenoid (see properties **C2** and **C4**) has two ends. Until recently, the only known examples of such surfaces were the plane, the helicoid and the catenoid. The helicoid is intimately related to the catenoid. (See FIGURE CH.)

2. Complete embedded minimal surfaces with finite total curvature

Finite topology and *finite total curvature*[6] are intimately related. If we take the example of the catenoid, we can get a feeling for this relationship. It can be

[5] A map $F\colon M \to N$ is proper if every divergent sequence in M is mapped into a divergent sequence in N.

[6] A surface is said to have *finite total curvature* if $\int |K|\, dA < \infty$.

The Catenoid/Helicoid

$$M = \mathbb{C} - \{0\}$$

$$g = z$$

$$\eta = cz^{-2}\,dz,$$

$$c = e^{i\theta}$$

Choose $p_0 = 1$,

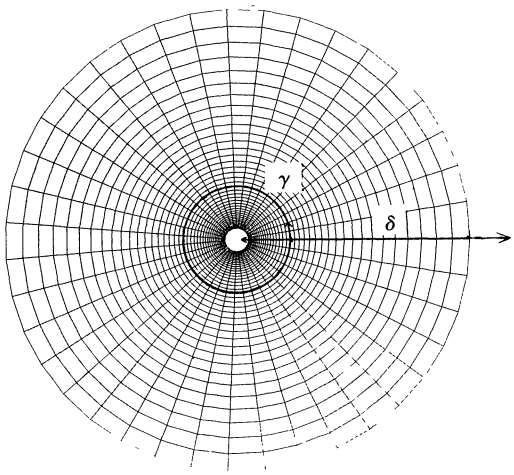
$$\Phi(z) = c\big((z^{-2} - 1), i(z^{-2} + 1), 2z^{-1}\big)\,dz,$$

$$X(z) = X_1, X_2, X_3$$

$$X_1 = -\mathrm{Re}\big(c(z^{-1} + z)\big),$$

$$X_2 = +\mathrm{Im}\big(c(z^{-1} - z)\big),$$

$$X_3 = 2(\cos\theta \ln|z| - \sin\theta \cdot \arg z)$$



$\lambda = 2(|z|^{-2} + 1)$ and is independent of θ.

Completeness:

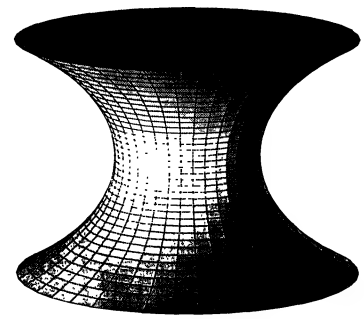
$$\int_1^{\infty} \lambda\,ds = 2\int_1^0 (t^{-2} + 1)\,|dt| = \infty$$

$$\int_1^0 \lambda\,ds = \int_1^0 (t^{-2} + 1)\,|dt| = \infty$$

The period vector

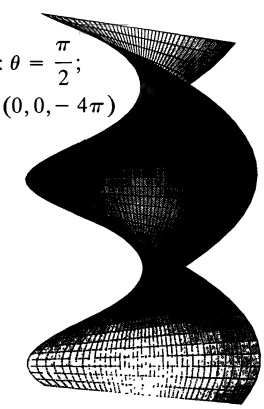
$$\vec{T}_\gamma = \int_\gamma \Phi = (0, 0, -4\pi \sin\theta)$$

Case 1. Catenoid: $\theta = 0$; $\vec{T}_\gamma = (0, 0, 0)$



Rays emanating from $\vec{0}$ go to meridians. Circles about $\vec{0}$ go to circles in planes $\{x_3 = \text{const.}\}$

Case 2. Helicoid: $\theta = \dfrac{\pi}{2}$;
$\vec{T}_\gamma = (0, 0, -4\pi)$



Rays go to horizontal straight lines
Circles go to helices
γ goes to the x_3-axis.

FIG. CH

guessed from the pictures, or directly computed from the formula for the catenoid, that the image under the Gauss normal map of each circle on the catenoid is a latitude circle on the sphere. Each meridian is covered, exactly once, so the entire sphere is covered, except for the poles. According to **C3**, we may identify the entire sphere with the catenoid by adding two points at infinity. Thus the meromorphic Gauss map is used to identify the compactified surface with the Riemann sphere, and the total curvature is just the area of the Gaussian image, counted algebraically; i.e., -4π. This relationship between conformal type and total curvature

was established in general by Osserman [**52, 53**] who proved that a complete minimal surface of finite total curvature:

K1) is conformally equivalent to a compact Riemann surface that has been punctured in a finite number of points and;

K2) has a Gauss normal mapping, defined and meromorphic on the punctured surface, which extends to a meromorphic function on the compactification. Also the holomorphic one-form associated to any coordinate function extends meromorphically to the conformal compactification.

Intuitively, the Gauss mapping provides a finite-to-one covering of the Riemann sphere. Because the curvature must go to zero quickly as one diverges, the principal curvatures, which are equal in absolute value on a minimal surface, must also go to zero quickly. This, together with the assumption of completeness, forces the normal to have a limiting value at the ends.

According to **K1**, finite total curvature implies finite topology. But the converse is *not* true, as is evident from the helicoid. The helicoid is topologically a once punctured sphere (i.e., it is simply-connected), but since it is periodic and not flat, it has infinite total curvature. (See FIGURE CH for a picture of the helicoid.)

Considering the fact that the helicoid was discovered in 1776 by Meusnier [**47**], it is surprising that it is at the present time *the only known example of a properly embedded minimal surface of finite topology but infinite total curvature.*

The authors have recently proved that *on a properly embedded minimal surface only two annular ends, at most, can have infinite total curvature: all the rest must have finite total curvature* [**33**]. This has interesting geometric consequences. It is well known [**59**] that:

K3) if an end of a properly embedded minimal surface has finite total curvature, it is asymptotic to either the end of a plane or the end of a catenoid. Moreover all these ends are parallel in the sense that their limit normals are parallel.

Specifically,

K3′) each end will eventually be a graph, over the exterior of a compact set in a plane (perpendicular to the limit normal at the end), of the form:

$$u(x_1, x_2) = c_0 + \alpha \ln \rho + \frac{c_1 x_1 + c_2 x_2}{\rho^2} + o(\rho^{-2}), \qquad \rho = |(x_1, x_2)|.$$

If $\alpha = 0$, the end is asymptotic to a plane, otherwise it is asymptotic to a catenoid.

If the end has finite total curvature, it is forced to be annular by **K3**. If we consider properly embedded minimal surfaces with finite topology, all the ends are annular by definition. The result mentioned above implies that, at most, two of the ends have infinite total curvature. Thus, according to **K3**, all the other ends look like the ends of planes or of catenoids.

Two important questions to be answered are the following. Given a properly embedded minimal surface:

Q2) If it is simply connected, must it be the plane or the helicoid?;

Q3) If it is not simply connected but has finite topology, must it have finite total curvature?

The question **Q3** is a strong generalization of **Q1**, as we will now explain. Consider first the general case of a complete embedded minimal surface with finite total curvature. We know from **K1** that it is a compact surface of finite genus, say

k, all of whose ends, say it has r of them, are annular. Assuming that our surface is not simply connected, each end possesses a distinguished nontrivial homotopy class, represented by a closed curve. This curve wraps once around the puncture point corresponding to that end. If we look for a length-minimizing curve in this homotopy class, we will always find one. The intuitive reason is this: the only thing that can happen to prevent this is that a sequence of curves, whose length goes to the minimum, will diverge in space. But by **K3**, this end ultimately looks like the end of a plane or of a catenoid, so it turns out that such a diverging sequence of loops will have to have *unbounded* length. It can be shown quite easily that the minimum-length curve, certainly a closed geodesic, exists [**10, 25**].

We observe that this closed geodesic is unique. If not, one of two things will happen: (i) either there are two closed geodesics in this homotopy class that do not intersect and bound an annulus; or (ii) there are two closed geodesics in the class that intersect, from which we can choose two subarcs that bound a simply connected region. We will apply the Gauss-Bonnet Formula, which states that

$$\int_R K\,dA = \int_{\partial R} \kappa_g\,ds + \sum \beta_i + 2\pi\chi(R) \tag{2.1}$$

on any compact subregion R of a Riemannian surface, where κ_g is the geodesic curvature of ∂R, the β_i are the exterior angles at the corners (if any) of ∂R, and $\chi(R)$ is the Euler characteristic of R. Since $K < 0$ almost everywhere on the surface, and our boundaries are geodesics:

$$\sum \beta_i + 2\pi\chi(R) < 0. \tag{2.2}$$

In case (i), there are no boundary angles and the geodesics must bound an annulus. Since the Euler characteristic of an annulus is equal to zero, we arrive at a contradiction in (2.2). In case (ii), two geodesic arcs form a geodesic 2-gon, bounding a disk-type region with Euler characteristic equal to one. Again a contradiction arises from (2.2), since the exterior angles are no bigger than π in absolute value.

By the same reasoning, it can be shown that closed-geodesic representatives of two *distinct* ends cannot cross. Let us assume for the moment that they do not meet tangentially. By removing each of the r distinct ends that they bound, we are left with a surface $\mathcal{S}$ of genus k, bounded by r closed geodesics. By the Gauss-Bonnet Formula (2.1), the total curvature of this surface is

$$2\pi\chi(\mathcal{S}) = 2\pi(2 - 2k - r). \tag{2.3}$$

Moreover, each closed geodesic *bounds an end* which is an annulus.

What is the total curvature of an embedded annular end that is bounded by each one of these geodesics? By **K3′** each of these ends is ultimately the graph, over a plane P orthogonal to the limit normal of the end, of a function that is asymptotic to a constant (planar end) or to $\alpha \cdot \ln(\rho) + \textit{constant}$, where ρ is distance from the origin in P, and α is a real constant. In both cases, for large ρ, the graph C_ρ over the circle of radius ρ is very close to a circle of radius ρ, and the graph itself for large ρ is very, very flat. Thus it is easy to believe and not hard to prove that the geodesic curvature of C_ρ on the surface is asymptotically equal to the geodesic curvature of the circle of radius ρ in P. But this is $\pm 2\pi$. Again we

apply the Gauss-Bonnet Formula (2.1) to the annulus bounded by the distinguished geodesic and C_ρ. Using the fact that the Euler characteristic of an annulus is zero and taking the limit as $\rho \to \infty$, we get

$$\int K dA = -2\pi, \tag{2.4}$$

for any embedded end of finite total curvature. Since we have r ends, they contribute $-2\pi r$ to the total curvature. It follows from this and (2.3) that

K4) on a complete embedded minimal surface of finite total curvature with embedded ends of genus k and r ends is exactly

$$\int K dA = 2\pi(2 - 2k - 2r) \tag{2.5}$$

(or, equivalently, the degree of the Gauss map is $r + k - 1$).

Remember we had to assume that the distinct closed geodesics corresponding to the ends did not touch. But what if they do? In general, if two geodesics meet at a point and are tangent there, they coincide. Thus if two closed-geodesic boundaries for distinct ends touch, they coincide, which implies that the entire surface is the union of two annuli, whose intersection is a closed geodesic. This means that the surface itself is a complete, embedded, minimal annulus, with total curvature equal to -4π, according to 2.4. (In particular, formula **K4** is true in general.) One can now argue in various ways to conclude that this minimal surface must be the catenoid. For example, we invoke a theorem of Osserman [**53, 54**], which states that

C5) The only complete nonsimply-connected minimal surface with total curvature -4π is the catenoid.

Or, we could quote a more recent theorem of Schoen [**59**]:

C6) The only complete embedded minimal surface of finite total curvature with two ends is the catenoid.

Or, finally we could appeal to a still more recent theorem of Lopez and Ros [**43**]:

C7) The only complete nonflat embedded minimal surface with finite total curvature and genus zero is the catenoid.

From this discussion, we see that once one knows that a complete embedded minimal annulus has finite total curvature it is the catenoid. In this sense **Q3**, considered as a conjecture, is a strong generalization of **Q1**, the Nitsche Conjecture.

2.1. New finite total curvature examples. The reader will notice that no examples of complete embedded minimal surfaces of finite topology have been mentioned, other than the plane, the helicoid and the catenoid. Questions **Q2** and **Q3** being open at the present time, any other examples would have to have finite total curvature. How would one construct such an example?

Notice that, according to **C6** and **C7**, any example will have to have genus $k \geq 1$ and at least 3 ends ($r \geq 3$).

Using properties **K1–K4**, combined with the Enneper-Weierstrass representation, one can follow this recipe to try to produce a new example:

S1) Choose a compact Riemann surface $\mathscr{R}$ of genus k, removing from it a finite number r of points. On $\mathscr{R}$, specify a meromorphic function g and a

holomorphic one-form η. To satisfy **K4** make sure that the degree of g is $k + r - 1$. To satisfy the second part of **K3**, g should be chosen to have zeros and poles at the ends.

S2) The Enneper-Weierstrass representation will produce a multivalued map of $\mathscr{R}$ into $\mathbb{R}^3$, which is a conformal minimal immersion of $\mathscr{R}$ minus the r punctures. Make sure the map is regular and complete. This can be done easily by looking at the poles and zeros of g and η. (See FIGURE EW.)

In general, the integration of the Enneper-Weierstrass one-forms around a nontrivial cycle produces a complex 3-vector, whose real part is what we refer to as the *period* associated to this cycle. Geometrically, the period is a translation in $\mathbb{R}^3$ produced by integration around the cycle. In general the Enneper-Weierstrass representation produces a periodic surface in $\mathbb{R}^3$ with the periods generating all the basic translations. In general these periods will be neither zero nor linearly independent (and there may be many of them). For example, the catenoid has no period corresponding to the cycle represented by the unit circle, while the helicoid has a period consisting of a vertical translation (see FIGURE CH). Usually the periods of an immersion result in a very messy surface in $\mathbb{R}^3$, not in general as nice as the helicoid. (In Sections 3 and 4, we will discuss periodic minimal surfaces constructed by trying to make one, two or three of the periods of an immersion both nonzero and linearly independent.

S3) Make sure that the mapping given by the Enneper-Weierstrass integrals has no residues at the punctures (the ends are well-defined) and *no periods* on any cycle on $\mathscr{R}$. This will mean that the Enneper-Weierstrass representation produces a finite total curvature minimal surface that has embedded, parallel ends. If the ends have distinct logarithmic growth (see **K3′**), this implies that the surface is embedded outside of a sufficiently large compact set of $\mathbb{R}^3$. Then:

S4) Show that the surface is embedded.

In 1981, C. Costa **[12, 13]** discovered an example that was conformally the square torus punctured in the half-periods. He did this by using elliptic functions to produce the Weierstrass analytical data; specifically $g = c\wp'$, and the one-form η was $\wp^{-1}\,dz$ where $\wp$ is the Weierstrass $\wp$-function. The surface was constructed to be complete and to be asymptotic, outside of a compact set to the union of a vertical catenoid ends and a horizontal plane separating them. (The value of c was computed algebraically to satisfy **S3**.) In particular, Costa's example was embedded outside of a compact set. Therefore, this example satisfies **S1**, **S2** and **S3**. In 1984, the authors **[31]** were able to prove that this surface was embedded.

We were subsequently able to construct examples of every positive genus $k \geq 2$ which were complete, embedded and possessed three ends **[29]**. Soon thereafter we proved that each of these surfaces, including the genus-one example, lies in a one-parameter family of embedded minimal surfaces of the same topological type **[30]**. In the genus-one case, Costa has recently shown that this family constitutes the only genus one complete minimal surfaces with three ends, all of which are embedded and parallel **[14]**.

These surfaces are defined by meromorphic data on compact Riemann surfaces, which data must be integrated to produce the coordinates of the surface, as specified in **S1** and **S2**. The residue problem **S3** must then be solved. The construction uses symmetry to annihilate the residues at the ends and to reduce the number of independent periods. For example, the surfaces that generalize

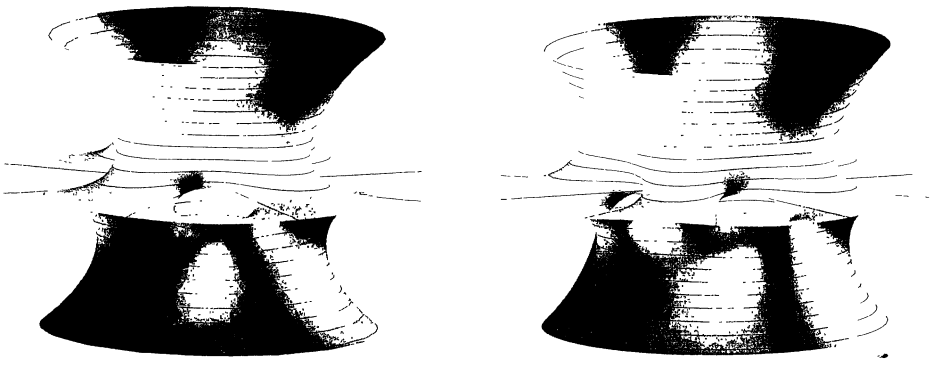


IMAGE 1. Embedded finite total curvature examples of genus 3 and 4 with 3 ends.

Costa's construction may be produced as follows:

G1) The underlying Riemann surface is the $(k+1)$-fold covering of the sphere given by

$$w^{k+1} = z^k(z^2 - 1),$$

punctured at the points $(1,0)$, $(-1,0)$ and (∞,∞). The Gauss map and one-form in FIGURE EW are given by

$$g = c/w, \qquad \eta = (z/w)^k\, dz,$$

where c is a real constant to be determined. The genus of this surface is k and it has 3 punctures. The degree of g is $k+2$ as required. Also, g takes on the values 0 or ∞ at the puncture points.

G2) Using the Enneper-Weierstrass representation (see FIGURE EW), we can compute that

$$X_3(z) = \text{constant} + c\left(\ln\left|\frac{1}{z-1}\right| - \ln\left|\frac{1}{z+1}\right|\right).$$

In this way these surfaces are similar to the catenoid; the top and bottom ends are of catenoid type. (See **K3'**.) This gives completeness at the ends at $(1,0)$ and $(-1,0)$. At (∞,∞), g is not one-to-one, as it would have to be on a catenoid type end. Therefore this end is flat. Observing that the first two components of the Enneper-Weierstrass integrand have poles at (∞,∞) one can show that the surface is complete at this end. The one-form η has a zero (order $= 2k$) on the punctured surface only at $(0,0)$, which is the only finite point where g has a pole (of order $= k$). This insures regularity (see FIGURE EW).

G3) The fact that there is so much symmetry simplifies the period computation enormously. This surface is constructed so that the image of any arc that projects onto the real axis and runs between ends is mapped by a branch of the immersion given by the Enneper-Weierstrass representation into a planar geodesic in a vertical plane. These planes must make an angle of $\pi j/(k+1)$ with the (x,z)-plane, $0 \le j \le k$. The image of a closed curve around any of the ends must therefore be invariant under reflection in each of these planes, which implies that

the residue at the end is orthogonal to each of these planes. Since there are more than one of them, this means that the residue is equal to zero.

For each k, the surface has $k + 1$ vertical planes of reflective symmetry. Any lift of a segment of the real axis is mapped into one of these planes. Thus we can divide the underlying Riemann surface into $2(k + 1)$ regions, corresponding to the $2(k + 1)$ distinct components of lifts of $\mathbb{C}-\mathbb{R}$. Furthermore, we can choose *one* closed curve in the z-plane whose lifts generate the homology of Riemann surface. Killing the period of the Enneper-Weierstrass integrand on this closed curve will kill all the periods on the surface. One can then vary the parameter in the definition of the g to kill this period: the proof provides an explicit closed formula for the value of c.

At this point, one knows that the surface exists and is minimal and has finite total curvature.

G4) At this point, we have a finite total curvature minimal surface that is embedded outside of a compact set. (The ends are embedded and disjoint.) A separate argument must be given to prove that the entire surface is embedded. Again symmetry is crucial. The surface can be decomposed into disjoint congruent pieces, each of which turns out to be a graph. This means the entire surface is embedded.

The authors also proved the existence of these surfaces by minimax methods [**32**]. This approach gives embeddedness automatically along with existence, but says nothing about uniqueness which can be proved by the analytic method. Pitts and Rubenstein have independently provided a minimax construction of these examples [**55**].

2.2. More finite total curvature examples. The authors and Michael Callahan [**9**] have constructed examples of complete embedded minimal surfaces of finite total curvature with *four* ends (two catenoidal, two flat). In principle, such examples exist with every even positive genus. Again, the equation of the underlying Riemann surface is deduced from symmetry considerations; the Gauss map and Weierstrass data are given in terms of roots of rational functions on a Riemann surface. In this respect, it is the same as the construction of the three-ended examples. However, there are in this construction *two* free variables (corresponding to the positioning of the middle flat ends and the height of the axial saddle point) and *two* independent periods. The proofs of existence and embeddedness of these surfaces depends heavily on machine computation. The computer is used to find the values of the free variables that annihilate the periods. Then, the degree of the period mapping is calculated, again with the aid of the computer. Finally, although the surface is symmetric, it is sufficiently complicated to make all the known methods for proving embeddedness inapplicable. The proof depends on a mathematical argument that, in essence, shows that the method used to produce the accurate coordinates of the surface, a program called MESH, devised and refined by J. Hoffman, is accurate [**8**]. Even this proof of embeddedness requires a machine computation, of a very simple sort, to complete it.

What is the general pattern? Karcher observed that these four-ended surfaces could probably be made more complicated by the insertion of a handle through their waist. This was carried out by M. Wohlgemuth, with the assistance of E. Boix in Amherst (Spring 1989). In this problem, the number of periods and free

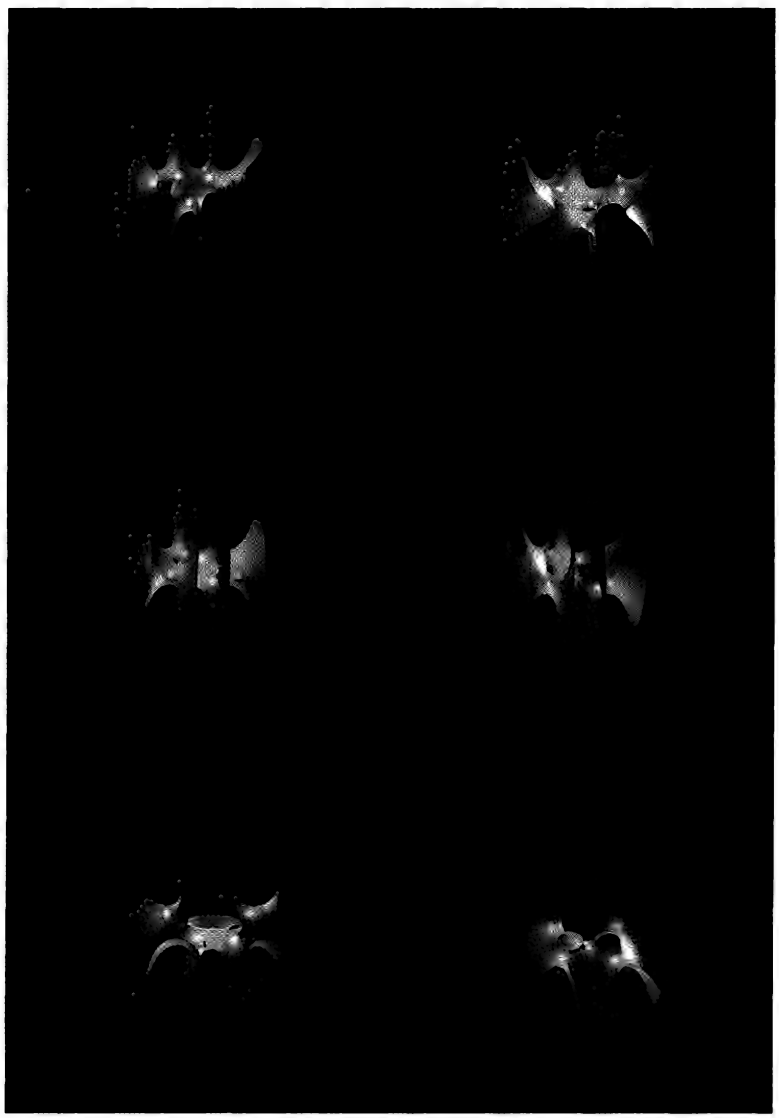


IMAGE 2. Four-ended examples. The top series shows the genus-two example of [9]. The bottom series shows a genus-three example of Wohlgemuth, produced by putting a handle on the example above.

parameters is three. Numerical and graphical searches were essential to the discovery of these surfaces, as was the use of MESH, which had been refined by J. Hoffman, working with Boix and Wohlgemuth, to deal with singularities and cuts in the domain of integration. The resulting surfaces have genus $3k$. This construction, combined with recent work of Wohlgemuth on constructing doubly-periodic surfaces by handle growing, indicates that there are probably a great number of different embedded minimal surfaces of finite total curvature and that if they are sufficiently symmetric, handles can be added to make them more complex without destroying the symmetry or adding any ends.

An interesting question is whether this handle adding construction can be carried out *ad infinitum* to create a limit surface Σ with an isolated singular point p, where the attaching handles accumulate. It is known [**24**] that a stable minimal variety or a minimal variety with finite topology cannot have an isolated "interior" singularity; i.e., such varieties extend smoothly across the potential singularity.

3. Minimal surfaces foliated by convex curves in parallel planes

3.1. Minimal annuli. Since it is a surface of revolution, the catenoid is fibred by circles in parallel planes. It is natural to try to classify all minimal annuli that are so constructed. We can argue in a straightforward fashion that a complete minimal annulus that is foliated by circles must in fact be the catenoid. Since any compact piece R bounded by parallel circles is an annulus (Euler characterize 0) each of whose boundary circles has curvature 2π in absolute value, we may use the Gauss-Bonnet Formula (2.1) to conclude that

$$\left|\int K\,dA\right| \leq 4\pi.$$

Thus the entire surface is an embedded complete annulus with finite total curvature (less than or equal to 4π in absolute value). We can use **C5** to conclude that this annulus is the catenoid.

In the next section, we will describe the classification of *compact* minimal annuli fibred by circles in parallel planes. First, we will describe some interesting results, old and new, about minimal annuli foliated by *convex* curves in parallel planes.

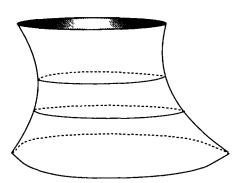


FIG. 2. A1) Minimal annuli bounded by convex curves in parallel planes are fibred by convex curves.

3.2. Schiffman's First and Second Theorems and extensions. Schiffman [**57**] proved two beautiful theorems on the geometry of a minimally immersed annulus A bounded by a pair Γ of smooth convex Jordan curves in parallel planes. Schiffman's first theorem states that

A1) A intersects each plane P, parallel to and between these planes, in a convex curve. In particular A is embedded.

Schiffman's second theorem states that

A2) when the convex curves Γ are circles, then $A \cap P$ is also a circle.

Thus, when Γ consists of circles, A is foliated by circles. We will now give a proof of **A1**. (For a proof of **A2** different from Schiffman's theorem see [**46**].)

Assume that the boundary curves of A are smooth convex curves in parallel planes. By uniformization, any annulus A is conformally equivalent to some annulus $A(r) = \{z \in \mathbb{C} | 1 \leq |z| \leq r\}$ for some unique $r = r(A) > 1$. After rotation and homothety, we may assume that A is the image of a conformal minimal immersion X: $A(r) \to \mathbb{R}^3$ with $X(|z| = 1) \subset P_0$ and $X(|z| = r) \subset P_{\ln(r)}$, where $P_t = \{x_3 = t\}$. The third coordinate function X_3: $A(r) \to \mathbb{R}$ is harmonic, and $\ln|z|$ is a harmonic function with the same boundary values. Note that both $\ln|z|$ and $X_3(z)$ are harmonic as functions on $A(r)$, because we are in conformal coordinates. We can therefore apply the maximum principle for harmonic functions to conclude that $X_3 = \ln|z|$. Since ∇X_3 is never zero on $A(r)$, it follows that the tangent plane to A is never horizontal. Thus $g(z)$, the stereographic projection of the Gauss map, is never 0 or ∞. This implies that $\phi = \arg g$ is well-defined, mod 2π. Notice that the horizontal projection of the normal to A at a point $p \in P_t \cap A$ is orthogonal to the plane curve $P_t \cap A$ at p and it is never zero. Hence ϕ is also the angle of the normal to $P_t \cap A$. If we parametrize $A(r)$ by polar coordinates (r, θ), then $P_{\ln(c)} \cap A$ is parametrized by $X(ce^{i\theta})$, $0 \leq \theta \leq 2\pi$. Observe that the convexity of $P_{\ln(c)} \cap A$ is equivalent to $(\partial/\partial\theta)\phi(ce^{i\theta})$ never changing sign. But g is holomorphic in z, and for any holomorphic function F, $(\partial/\partial\theta)(\arg F)$ is harmonic. Since both boundary curves of A have the same orientation, the harmonic function $(\partial/\partial\theta)\phi$ has a fixed sign on ∂A. By the maximum principle, it is never zero in the interior, and we are done.

Recently Meeks and White have proven a theorem that augments Schiffman's first theorem [**46**]. The proof of this theorem depends on stability properties of compact minimal surfaces.

We shall call a compact minimal surface M *stable* if, with respect to any nontrivial normal variation fixing the boundary, the second derivative of area is positive. If the second derivative of area is negative for some variation, then M is called *unstable*. If M is neither stable nor unstable, we will call it *almost-stable*.

A3) If Γ is a pair of smooth, convex Jordan curves in distinct parallel planes, then exactly one of the following holds:

1. Γ is not the boundary of any connected compact minimal surface, with or without branch points;
2. Γ is the boundary of exactly one minimal annulus and this annulus is almost-stable. In this case, Γ bounds on other connected compact branched minimal surface;
3. Γ is the boundary of exactly two minimal annuli; one stable and one unstable.

What other compact minimal surfaces may have Γ as their boundary in case 3 is unknown. That is:

Q4) Given two convex curves in parallel planes, can they bound a minimal surface of genus greater than zero?

In certain cases it is known that every connected branched minimal surface with boundary Γ described in **A3** is actually an annulus. For example, R. Schoen [**59**] proved that when Γ is contained in parallel horizontal planes and is invariant under reflection in two vertical planes, then every branched minimal surface with

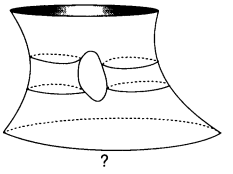


FIG. 3. It is not known whether convex curves in parallel planes can bound a surface of genus greater than zero.

boundary Γ is actually an annulus. (This was used in the proof of **C6**.) Thus, in certain cases, **A3** shows that Γ is the boundary of 0, 1, or 2 minimal annuli and no other branched minimal surfaces.

As a simple corollary to **A3**, we can prove Schiffman's Symmetry Theorem.

A4) A minimal annulus bounded by convex curves in parallel planes has the symmetries of its boundary.

Suppose A is such an annulus and S: $\mathbb{R}^3 \to \mathbb{R}^3$ is a symmetry of the boundary of A. Then $S(A)$ is another minimal annulus with the same boundary and $S(A)$ has the same stability properties as A. Hence by **A3**, $S(A)$ must equal A, which proves **A4**. Statement **A4** was proved by Schiffman using a different method [**58**].

The proof of **A3** uses some new techniques that come from global analysis. Consider the space $\mathscr{M}$ of smooth embedded minimal annuli in $\mathbb{R}^3$ with boundary curves in two fixed parallel planes. Let p: $\mathscr{M} \to \mathscr{C}$ be the associated projection onto the space of pairs of smooth Jordan curves that are contained in the parallel planes. A basic result (see [**4, 6, 63, 64, 65,**]) is that $\mathscr{M}$ and $\mathscr{C}$ are smooth infinite dimensional manifolds and p is a proper Fredholm map of index zero. In this setting one can apply the classical results of global analysis, such as the Sard-Smale Transversality Theorem [**62**], to analyze the geometry of minimal annuli with boundary $\Gamma \in \mathscr{C}$.

In order to explain how these techniques might apply to prove **A3**, we examine a special case. Suppose Γ is a pair of coaxial circles in parallel planes P_1, P_2 and $\mathscr{M}$ is the corresponding space of minimal annuli for P_1, P_2. Since a minimal annulus A with $\partial A = \Gamma$ must have all the symmetry of its boundary, A must be a surface of revolution. This means A must be a catenoid, but notice that this does not mean A is the only catenoid with boundary Γ. One can check directly from the equation of a catenoid (1.5) that Γ is the boundary of n catenoids, where $n = 0$, 1, or 2. If $n = 1$, then the unique catenoid with boundary Γ is almost-stable. If $n = 2$, then Γ is the boundary of one stable and one unstable catenoid.

Now shrink the circles in Γ through a one-parameter family $\Gamma(t)$, $0 \le t \le 1$, of coaxial circles in $P_1 \cup P_2$ so that $\Gamma(0) = \Gamma$ and the circles in $\Gamma(1)$ are so small that $\Gamma(1)$ does not bound a catenoid. Consider $\Gamma(t)$ to be a path in $\mathscr{C}$ and we get the picture described in FIGURE 4.

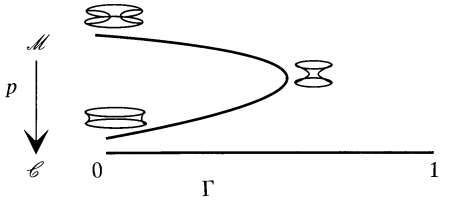


FIG. 4

That the above picture generalizes to a pair of convex curves, $\Gamma \subset P_1 \cup P_2$ is used to prove **A3**. One shrinks Γ through a one-parameter family $\Gamma(t)$, $0 \leq t \leq 1$, of smaller and smaller convex curves in $P_1 \cup P_2$ so that $\Gamma(0) = \Gamma$ and $\Gamma(1)$ bounds no minimal annulus. Then $\Delta = p^{-1}(\Gamma[0, 1]) \subset \mathcal{M}$ is either empty, consists of a unique minimal annulus with boundary Γ, or is a compact arc. In the last case, the projection of this arc onto $\Gamma[0, 1]$ has one fold point Σ corresponding to an almost-stable minimal annulus and $\Delta - \{\Sigma\}$ consists of two components, each consisting entirely of stable or unstable annuli.

3.3. Minimal surfaces foliated by circles and lines in parallel planes: Riemann's examples. Recall that the catenoid is a complete minimal surface that is foliated by circles in parallel planes. By **A2**, a minimal annulus with circle boundaries on parallel planes is foliated by circles in parallel planes. Given two circles, in parallel planes, whose centers are sufficiently close together, it is possible to find a minimal annulus spanning them[7]. Thus, there exist compact minimal annuli, other than portions of a catenoid, that are foliated by circles in parallel planes. It is a surprising fact that these circle-foliated minimal annuli are subsets of complete, embedded, periodic minimal surfaces. (This was proved by Enneper [**19**]. See also [**51**], Sections 94–96 and [**35**].) These minimal surfaces, foliated by circles and lines in parallel planes, were discovered by Riemann [**56**], who showed that they could be described in terms of elliptic functions on rectangular tori. We will now give a geometric argument that determines the analytic description of these circle-foliated surfaces. Along the way we will prove that they are subsets of complete, embedded, periodic minimal surfaces.

First we recall the Schwarz reflection principle for minimal surfaces ([**60**], also see [**41**] for the proof). This principle states that

R1) A minimal surface M with a line in its boundary may be extended to a larger minimal surface by reflection across that line, and

R2) A minimal surface M, which meets a plane orthogonally, may be extended to a larger minimal surface by reflection in that plane.

By reflection in a line, we mean rotation by π about the line. Minimal surfaces are analytic surfaces, as can be deduced from the Enneper-Weierstrass representation or from the fact that their coordinate functions are harmonic. It follows from this fact and **R1** and **R2** that if a minimal surface contains a line (resp., a planar geodesic) in its interior that surface is invariant under rotation about that line (resp. reflection in that plane). The Schwarz Reflection Principles for minimal surfaces follow from the Reflection Principle for harmonic and analytic functions.

In 1951 H. Lewy [**42**] gave an important generalization of the Schwarz reflection principle. He proved that

R3) a minimal surface with analytic boundary curves, is contained in the interior of a larger minimal surface.

Suppose now that A is a compact minimal annulus that is foliated by circles in parallel planes. We first show that A is contained in a larger compact annulus $\tilde{A}$

[7] You can create such a surface by fashioning two circles out of wire and dipping them in a soap solution. For advice about details and more discussion of soap films as minimal surfaces, see [**7, 15, 27, 51**]. The general problem of finding a minimal surface with a given contour as boundary is called Plateau's Problem after the Belgian physicist J. A. F. Plateau. This problem is a central one in the study of minimal surfaces, which we only mention in passing in this article.

foliated by circles in parallel planes. Since the boundary of A is analytic and A is everywhere transverse to horizontal planes, Lewy's theorem shows that A is contained in the interior of another compact minimal annulus $\tilde{A}$, whose boundary consists of analytic curves in planes parallel to the planes containing ∂A. Given an analytic surface and a family of parallel planes, parametrized by their distance from a fixed plane, that intersect the surface transversally, the intersection of each plane with the surface is an analytic curve. The family of such curves depends analytically on the parameter. In the case we are considering, the analytic one-parameter family consists of circles for some nonempty closed interval corresponding to the original annulus A. Hence $\tilde{A}$ must be fibred by circles in parallel planes. In particular, $\tilde{A}$ is bounded by circles.

Let F be the maximal open annulus that contains A and is fibred by closed curves in parallel planes. It follows that F is fibred by circles. Any pair of these circles is symmetric with respect to some vertical plane P. Either that plane is unique, or that pair of circles has centers lying on a vertical line, in which case every rotation about that axis is a symmetry. From **A4**, it follows that the annulus bounded by these circles either possesses a unique plane of vertical symmetry or is a surface of rotation. In the latter case, we can conclude immediately that we have a piece of the catenoid and that F is a catenoid. In the former case, we can conclude that every compact subannulus of F bounded by circles has a unique fixed plane of symmetry P and hence F has the same property.

$P \cap F$ consists of two arcs σ_1, σ_2, each a planar geodesic line of curvature, meeting each circle in F orthogonally at antipodal points. It follows that the tangent vectors to the circles along these arcs are principal directions. Without loss of generality, we may presume that P is the (x, z)-plane. Since the surface is minimal, the principal curvatures at each point are equal in absolute value, with opposite signs. But the curvature vectors of the circles are horizontal and never zero and the normal vector to F is never vertical. Thus it follows that the curvature vectors to σ_1 and σ_2, which are vectors in the plane P, are nonzero and oppositely oriented. In particular, each σ_i is a strictly convex curve in P. It is natural to parametrize σ_i by t, letting $\sigma_i(t) = \sigma_i \cap \{z = t\}$. Let $r(t) = \frac{1}{2}|\sigma_1(t) - \sigma_2(t)|$ be the radius of the circle $F \cap \{z = t\}$.

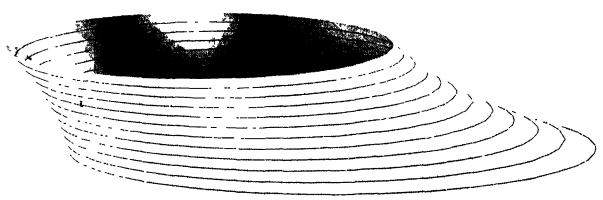


FIG. 5. A minimal annulus foliated by circles in parallel planes.

Observe that by the Gauss-Bonnet formula (2.1) the total curvature of every compact subannulus bounded by circles is not greater than 4π in absolute value. Thus the total curvature of F is at most 4π in absolute value. If F is regular and complete, then by **C5** it must be a catenoid.

If F is not complete, then one or both of the convex curves σ_i must have a finite endpoint. If they both terminate at distinct endpoints at the same level $\{z = t_0\}$,

then $\bar{F}$ is bounded by a circle at that level. But by **R3**, this means that F lies in the interior of a larger minimal surface containing that circle in its interior. Such a surface is still fibred by horizontal curves near the circle at level $\{z = t_0\}$, and hence, as observed above, by circles. This contradicts the assumption that F is the maximal such minimal annulus.

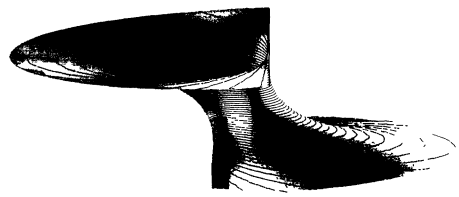


FIG. 6. The maximal annulus $\bar{F}$.

Suppose that $\sigma_1(t)$ and $\sigma_2(t)$ both terminate at the same point at height t_0. Without loss of generality, we may assume that F lies in the halfspace $\{z < t_0\}$. We will show that this is impossible by computing the flux, in the direction of the vertical unit vector field, for the minimal annulus $F \cap \{t_2 < z < t_1\}$, where $t_1 < t_0$ and t_2 is chosen large enough so that $F \cap \{z = t_2\} \neq \emptyset$. The flux through the boundary circle at level t_2 is nonzero. The flux through the boundary circle at level t_1 is bounded above in absolute value by $2\pi r(t_1)$, which goes to zero as $t_1 \to t_0$. According to the flux formula (1.3) these two quantities must be equal in absolute value, a contradiction.

We may conclude that if one of the curves $\sigma_i(t)$ terminates in a point at level t_0, then the other $\sigma_i(t)$ diverges to infinity as $t \to t_0$. This means that $\bar{F}$ has a boundary curve at height t_0 consisting of a horizontal line. Suppose that this is the only component of ∂F. Then, according to **R2**, we may extend $\bar{F}$ by rotation about this line, producing a complete, embedded genus-0 minimal surface with 3 ends. We observed above that F had total curvature at most 4π in absolute value, so this extended surface has total curvature of at most 8π in absolute value. By **C7**, we conclude that this surface must be the catenoid, which is impossible because it contains a line (and has three ends!). Thus there must be another boundary component of $\bar{F}$, and the arguments above imply that it, too, is a straight line.

Finally, we consider the case that σ_1 and σ_2 have no endpoints at all. In this case, F is a complete, embedded minimal annulus with total curvature at most 4π, in absolute value, and so must be the catenoid. We have thus determined that either $\bar{F}$ is the catenoid or it is an annulus bounded by two lines L_1, L_2, each orthogonal to the plane of symmetry P, lying in parallel planes.

From now on, we will assume $\bar{F}$ is not the catenoid. The function $r(t)$ is thus defined on an interval and diverges to infinity at the endpoints. We may therefore find two distinct values of t say $t_1 < t_2$, where $r(t_1) = r(t_2)$. The corresponding circles at heights t_1 and t_2 are invariant under rotation about a line l at height $(t_1 + t_2)/2$ which is perpendicular to P and placed halfway between their centers.

By **A4**, the annulus $F \cap \{t_1 \leq z \leq t_2\}$ is invariant under rotation about l, which implies that F has the same property. This rotation interchanges σ_1 and σ_2, so they are congruent. In particular, each one has one endpoint on one of the lines L_i.

By **R1**, $\bar{F}$ can be extended by rotations, R_1, R_2 across line boundaries L_1, L_2, respectively. Let G be the group of symmetries generated by R_1 and R_2 and let $\mathscr{R}$ denote the G-orbit of $\bar{F}$. Note that $\mathscr{R}$ is invariant under the translation $T = R_1 \circ R_2$ and $\mathscr{R}/T$ is topologically a compact torus $\overline{\mathscr{R}}$ punctured in two points p_1, p_2, corresponding to the points at infinity on the two lines in $\bar{F} \cup R_1(\bar{F}) = \mathscr{R}/T$.

To determine the conformal type of $\mathscr{R}/T$, we may use the fact that it is highly symmetric. Reflection through P produces an orientation-reversing automorphism that commutes with the orientation-reversing automorphisms corresponding to rotation about L_1 or L_2. Straightforward arguments can be given to show that $\overline{\mathscr{R}}/T$ is conformally diffeomorphic to $\mathbb{C}/\Lambda(\lambda)$, where $\Lambda(\lambda) = \{m\lambda + ni | n, m \in \mathbb{Z}\}$, for some $\lambda > 0$, i.e., it is a rectangular torus. Moreover, reflection in P will correspond to complex conjugation (fixing the curves t and $t + i/2$, $0 \leq t \leq \lambda$, which correspond to σ_1 and σ_2) and rotation about L_2 corresponds to reflection in the imaginary axis. We may translate so that the puncture, p_1, corresponding to the end into which L_2 diverges, is at $\vec{0}$ (assuming L_i crosses σ_i.) This will force rotation about L_2 to correspond to reflection in the line $\lambda/2 + it$, $0 \leq t \leq 1$. Since the line L_2 crosses σ_2 but not σ_1, it must be that the other end, p_2, is at the point $w_3 = (\lambda + i)/2$.

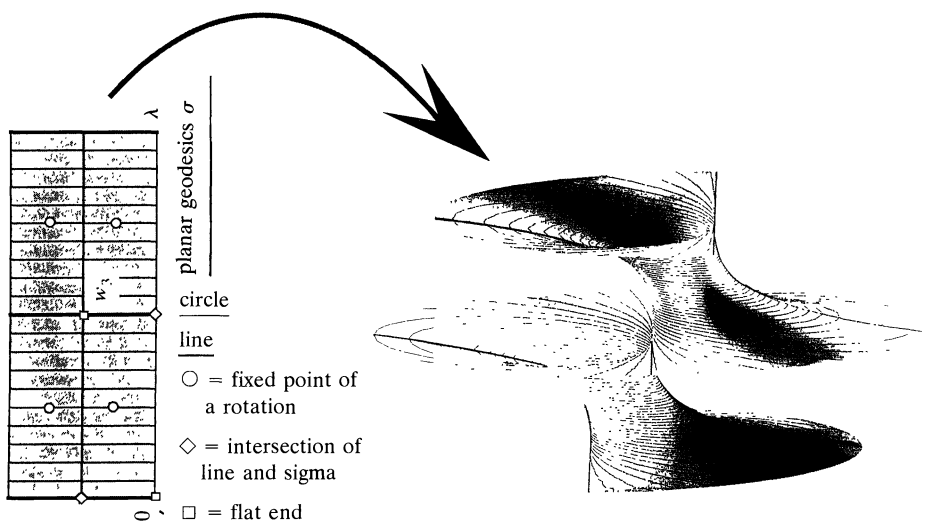


FIG. 7. 'Riemann's example parametrized by a rectangular torus.

Thus we have determined the underlying Riemann surface structure of $\mathscr{R}/T$. Furthermore, we know that the total curvature of F, computed as the limit of minimal annuli via the Gauss-Bonnet formula must be -4π. Hence the total curvature of $\mathscr{R}/T$ is -8π, which means that the Gauss map has degree 2. Furthermore, the Gauss mapping is vertical at $\vec{0}$ and w_3; without loss of generality we may assume that $N(\vec{0}) = (0, 0, 1)$ and $N(w_3) = (0, 0, -1)$. In terms of g, the stereographic projection of N, $g(\vec{0}) = \infty$ and $g(w_3) = 0$. Since the ends at $\vec{0}$ and w_3 are flat, not catenoidal, the pole at $\vec{0}$ and the zero at w_3 must have orders at least two. But the degree of g is 2, which implies that g has a pole of order 2 at $\vec{0}$, a zero of order 2 at w_3, and no other zeroes or poles. From the basic theory of

elliptic functions, we can conclude that $g = c[\wp(z) - \wp(w_3)]$, where $\wp$ is the Weierstrass $\wp$-function. Moreover, the third coordinate function x_3 is constant on the lines ti, and $\lambda/2 + ti$, $0 < t < 1$. Again we can argue that $(d/dz)(x_3)$ must be equal to a constant. By using the Enneper-Weierstrass representation formula, we see that $\eta = dz/g$. Rotation about the line l fixes two points of $\mathscr{R}/T$, and interchanges the lines L_1 and L_2. It follows that we may choose these points to be $w_3/2$ and $3w_3/2$. At these points the normal to $\mathscr{R}/T$ is horizontal, so g must be unitary. We adjust c, if necessary, to make this happen. We omit the simple argument which shows that c may be chosen to be real.

4. Embedded periodic minimal surfaces and the period problem

The surfaces of Riemann $\mathscr{R}$ are foliated by circles and straight lines and have the appearance of an infinite staircase. In fact, after normalizing $\mathscr{R}$ to have the (x,z)-plane as a symmetry plane and the horizontal lines at integer heights, $\mathscr{R}$ can be approximated by the union of parallel planes at integer heights joined by ascending tubes. Thus we see that $\mathscr{R}$ has an infinite number of annular ends, each of which is asymptotic to a horizontal plane at an integer height. In particular, $\mathscr{R}$ has genus zero and an infinite number of ends. Moreover there are infinitely many different such surfaces, one for each rectangular torus.

M. Callahan and the authors [**11**] have found new examples of periodic minimal surfaces $\mathscr{M}$ that have much in common with $\mathscr{R}$. Qualitatively, they may be described as an infinite number of planes connected not by genus-zero tubes but by genus-k minimal surfaces, which approximate compact pieces of the surfaces described in **G1–G4**. (See IMAGE 1.) These surfaces $\mathscr{M}$ are invariant under a translation **T** and $\mathscr{M}/\mathbf{T}$ has genus $2k + 1$. Recently, new periodic examples $\mathscr{N}$ have been found by Hoffman and Wohlgemuth [**34**], with the assistance of E. Boix. They resemble the surfaces $\mathscr{M}$ with a handle, parallel to the planes, grown through the waist. The quotient by translation of these surfaces can have genus $3k + 1$ for any $k \geq 1$.

In the construction of the Riemann examples $\mathscr{R}$, there was no need to annihiliate a period in order to produce a well-defined surface; the surfaces were determined by geometric considerations and lie in a one-parameter family. In the Enneper-Weierstrass representation, the only period was the one that generated the desired translation. For the examples $\mathscr{M}$, the situation is quite different. The examples of a fixed genus are isolated. The Enneper-Weierstrass representation, derived from geometric considerations, has a free variable that must be adjusted to annihilate a period of the mapping. In [**11**], it is proved that this is possible. For the examples $\mathscr{N}$, the situation is more complicated still; there are two unwanted periods that must be annihilated and two free variables in the Enneper-Weierstrass representation. The periods are calculated on a computer.

Graphics plays an important role in understanding the geometry of these complex surfaces as well as in checking that the period computations are accurate. The proof of existence of the surfaces $\mathscr{N}$ requires a degree theory argument, which in turn is based on a machine computation. This means that the existence proof (analog of **S1–S3**) must be done genus-by-genus.

This situation is unsatisfactory. On the one hand, it now appears that we can make these surfaces more and more complicated by growing more and more

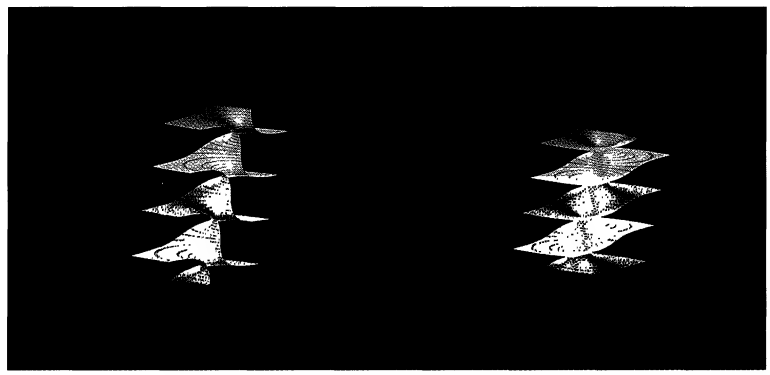


IMAGE 3. The Riemann example (right) and half of the Riemann example (cut by the plane P).

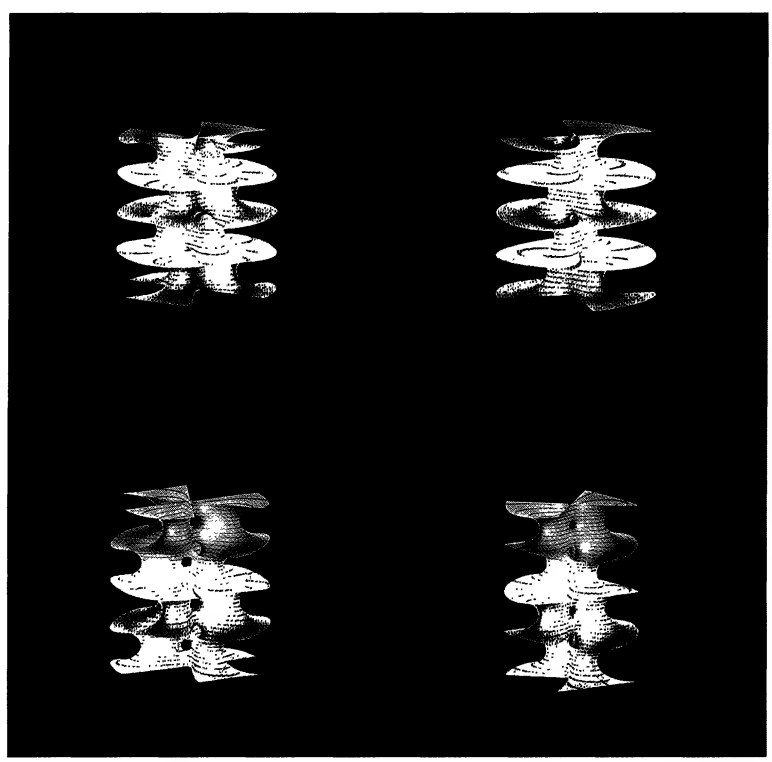


IMAGE 4. The periodic examples of [**11**] and their more complex cousins, produced by handle-adding, in [**34**].

handles and solving more and more complicated period problems by computer. Current experience with deriving the equations and using graphics tools to help solve the period problem indicates that these more complex surfaces are there. At present we do not understand the general behavior of the period mapping, nor do we have any general method of proving that the period problem can be solved in the setting of periodic minimal surfaces with a finite total curvature quotient, or even in the finite total curvature setting of Section 2.

Computation and graphics have indicated that these constructions can be carried out in situations where, beforehand, it would have been very hard to even guess that the surfaces existed. However, a solution to the problem of periods, even for a restricted class of examples, is a theoretical problem that is far from being solved.

Fortunately, we do know a good deal about the qualitative behavior of *properly embedded periodic minimal surfaces with more than one topological end*. If the symmetry group is not discrete it is the catenoid. Otherwise, [**10**]

P1) A properly embedded periodic minimal surface M with more than *one* topological end and discrete symmetry group:

i) has an infinite number of ends;
ii) is invariant under a screw motion S;
iii) all the annular ends of M are flat ends;
iv) the total curvature of M/S is finite if and only if it has finite topology.

(A *screw motion* is the composition of a translation and a rotation along an axis of the translation.) It is thus the case that the analog of **Q3** has an affirmative answer in this context. See also **P3** below.

As a relatively easy consequence of these and companion results, it can be shown [**10**] that:

T1) A properly embedded minimal surface that is doubly-periodic must necessarily have *one* topological end.

Certainly, properly embedded triply-periodic minimal surfaces have one topological end. There are also singly-periodic examples with one end, Scherk's one-periodic surface for example (see IMAGE 5). Property **P1** iv) generalizes to a properly embedded periodic minimal surface M invariant under a discrete group G, in the case M/G has finite topology [**44, 45**]:

P2) Suppose M is a properly embedded minimal surface in a nonsimply-connected flat three-manifold. Then M has finite topology if and only if it has finite total curvature.

P3) If M is a complete minimally immersed surface of finite total curvature in a flat three-manifold, then M is conformally diffeomorphic to a closed Riemann surface $\overline{M}$ punctured in a finite number of points. Furthermore, M can be defined analytically in terms of two meromorphic one-forms on $\overline{M}$.

These results are obtained by studying the asymptotic geometry of the quotient surface in the flat three-manifold $\mathbb{R}^3/G$, and were motivated by earlier constructions of Karcher [**36**] using the analytic technique described in **P3**.

All known examples, M, of properly embedded minimal surfaces in $\mathbb{R}^3$ are actually invariant under a discrete group G of symmetries that act freely on $\mathbb{R}^3$, and possess the property that the quotient surface M/G has finite total curvature. (When M itself has finite total curvature, the group G is the trivial group. In this

case the representation is described in **K1–K4**.) It therefore follows from **P3** that all known M are determined by a generalized Enneper-Weierstrass representation, using two meromorphic one-forms on a closed Riemann surface. We believe that there exist properly embedded minimal surfaces in $\mathbb{R}^3$ that cannot be defined in terms of meromorphic forms on a closed Riemann surface but we lack a method for constructing them.

As we have observed in **Q1**, it is unknown whether or not the plane and the helicoid are the only properly embedded simply connected minimal surfaces in $\mathbb{R}^3$. However, using **P3** and **P4**, it has been proved that the helicoid and the plane are the only properly embedded simply connected minimal surfaces in $\mathbb{R}^3$ with an infinite symmetry group [**44**].

5. Conclusion

In this article, we have tried to give a survey of some recent progress in the theory of minimal surfaces, while at the same time illustrating some of the underlying techniques and stressing the connection with the classical point of view. We wish to conclude by discussing briefly two topics—one topological, the other structural.

5.1. The topology of proper periodic minimal embeddings in R^3. In the previous section, properly embedded minimal surfaces with an infinite symmetry group and more than one end were discussed. Except for the catenoid, they all had an infinite number of ends and were singly-periodic with a discrete symmetry group. Any other periodic minimal surface has one topological end and, of course, infinite genus.

From the viewpoint of a topologist, it is reasonable to ask whether two properly embedded minimal surfaces of infinite genus and one end are ambiently isotopic. In other words, can one such surface be deformed through embedded surfaces to the other surface by a one-parameter family of diffeomorphisms of $\mathbb{R}^3$? The answer to this question is yes [**23**].

T2) If M_1, M_2 are two properly embedded minimal surfaces of infinite genus and one end, then there is a homeomorphism $h\colon \mathbb{R}^3 \to \mathbb{R}^3$ such that $h(M_1) = M_2$.

Even for particular examples of minimal surfaces satisfying the hypotheses of the above theorem, it is not obvious that the surfaces are ambiently isotopic. For example, consider Scherk's singly-periodic and doubly-periodic minimal surfaces in IMAGE 5. Each of these surfaces has infinite genus and one end, but it is difficult to imagine a homeomorphism of $\mathbb{R}^3$ taking one of these surfaces to the other.

5.2. Periodic interfaces in materials as a source of new minimal surfaces. In compound materials, bonded polymers for example, equilibrium structures are formed, which tend to minimize the area of the interface between the materials. It is therefore not surprising that minimal surfaces and surfaces of constant mean curvature are reasonable models of the resulting interface. More surprising is the fact that in recent years, new periodic embedded minimal and constant mean curvature interfaces have been discovered by material scientists [**1, 2, 20, 21, 58, 61**], and these discoveries have led to the discoveries of other new surfaces and to interesting theoretical work and collaboration [**3, 38, 50**].

We sketch briefly just one instance of this new input into the subject. In [**21**], Fischer and Koch, crystallographers, took a systematic approach to the construc-

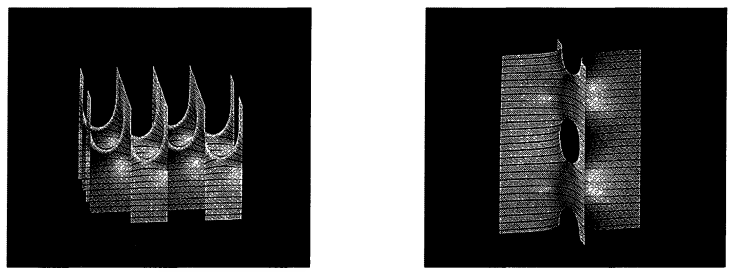


IMAGE 5. Scherk's singly and doubly-periodic surfaces.



IMAGE 6. Singly-periodic examples related to triply-periodic surfaces with noncubic symmetry.

tion of triply periodic embedded minimal surfaces. They studied the subgroups of the crystallographic groups of $\mathbb{R}^3$, looking for those that possessed a fundamental domain that contained extremal polygons, which were reasonable candidates for boundaries of fundamental pieces of triply-periodic minimal surfaces. One of the new surfaces they discovered is illustrated in IMAGE 6.

On viewing a computer image of this surface, produced by O. Wohlrahb in Bonn, it was observed [**28**] that a similar construction could be used to create families of singly-periodic minimal surfaces, with flat ends (like Scherk's single-periodic surface). The quotient of these new examples, by translation, are conformally rhombic tori. It is curious to note that the Enneper-Weierstrass representation for these surfaces was determined by using the images of the surface as a guide, a complete reversal of the procedures described in the previous two sections.

Acknowledgments. The authors wish to thank James T. Hoffman for producing the color computer graphics that illustrate this article, as well as many of the monochrome computer graphics of surfaces. We also wish to thank Eric Boix for scientific programming relating to the determination of some of the new minimal surfaces discussed in this article, as well as for assistance in producing the images. Thanks to James Riordan for computing the coordinates of the singly-periodic surfaces derived from the crystallographers' examples. We are grateful to Rob Kusner for a critical reading of an earlier version

of this text. All this collaboration used the facilities of GANG, the Center for Geometry, Analysis, Numerics and Graphics at the University of Massachusetts. We also wish to thank Rick Schoen for directing our attention to [**19**], and J. C. C. Nitsche for useful historical comments.

REFERENCES

1. D. Anderson, Studies in the microstructure of microemulsions, Ph.D. thesis, University of Minnesota, June 1986.
2. D. Anderson, C. Henke, D. Hoffman, and E. L. Thomas, Periodic area-minimizing surfaces in block copolymers, *Nature*, 334 (1989) 598–601.
3. D. M. Anderson, H. T. Davis, J. C. C. Nitsche, and L. E. Scriven, Periodic surfaces of prescribed mean curvature, to appear in *Advances in Chem. Phys.*
4. M. Anderson, Curvature estimates for minimal surfaces in 3-manifolds, *Ann. Scient. Éc. Norm. Sup.*, 18 (1985) 89–105.
5. L. Barbosa and A. G. Colares, Minimal Surfaces in $\mathbb{R}^3$, number 1195 in Lecture Notes in Mathematics, Springer-Verlag, 1986.
6. R. Bohme and A. J. Tromba, The index theorem for classical minimal surfaces, *Annals of Math.*, 113 (1981) 447–499.
7. C. V. Boys, Soap Bubbles: Their Colours and the Forces Which Mold Them, Dover Publications, New York, 1959.
8. M. Callahan, D. Hoffman, and J. Hoffman, Computer graphics tools for the study of minimal surfaces. *Communications of the ACM*, 31 (1988) 648–661.
9. M. Callahan, D. Hoffman, and W. H. Meeks III, Embedded minimal surfaces with four ends, in preparation.
10. _____, The structure of singly-periodic minimal surfaces, to appear in *Inventiones Math.*
11. _____, Embedded minimal surfaces with an infinite number of ends, *Inventiones Math.*, 96 (1989) 459–505.
12. C. Costa, Imersöes minimas en $\mathbb{R}^3$ de gênero un e curvatura total finita. Ph.D. thesis, IMPA, Rio de Janiero, Brasil, 1982.
13. _____, Example of a complete minimal immersion in $\mathbb{R}^3$ of genus one and three embedded ends, *Bull. Soc. Bras. Mat.*, 15 (1984) 47–54.
14. _____, Classification of complete minimal surfaces in $\mathbb{R}^3$ with total curnature 12π, preprint IMPA 1989.
15. R. Courant, Soap film experiments with minimal surfaces, *Amer. Math. Monthly*, 47 (1940) 167–174.
16. _____, Dirichlet's Principle, Conformal Mapping and Minimal Surfaces, Interscience Publishers, Inc., New York, 1950.
17. G. Darboux, Leçons sur la théorie générale des surfaces et les applications géometriques du calcul infinitésimal, Gauthier-Villars, Paris, 1st part, 2nd edition, 1914.
18. M. do Carmo and C. K. Peng, Stable minimal surfaces in $\mathbb{R}^3$ are planes, *Bulletin of the AMS*, 1 (1979) 903–906.
19. A. Enneper, Die cyklischen flachen, *Z. Math. und Phys.*, 14 (1869) 393–421.
20. W. Fischer and E. Koch, On 3-periodic minimal surfaces, *Zeitschrift für Kristallographie*, 179 (1987) 31–52.
21. _____, On 3-periodic minimal surfaces with noncubic symmetry, *Zeitschrift für Kristallographie*, 183 (1988) 129–152.
22. D. Fischer-Colbrie and R. Schoen, The structure of complete stable minimal surfaces in 3-manifolds of nonnegative scalar curvature, *Comm. Pure Appl. Math.*, 33 (1980) 199–211.
23. C. Frohman and W. H. Meeks III, The topological uniqueness of complete one-ended minimal surfaces and Heegaard surfaces in $\mathbb{R}^3$, preprint.
24. R. Gulliver and H. B. Lawson, The structure of minimal hypersurfaces near a singularity, *Proc. Symp. Pure Math.*, 14 (1986) 213–237.
25. J.-J. Hadamard, Les surfaces à courbures opposées et leurs lignes géodésiques, *Journal de Mathématiques Pures et Appliquées*, 4 (1898) 27–73.
26. S. Hildebrandt, The calculus of variations today, *Mathematical Intelligencer*, 11 (1989) 50–60.
27. S. Hildebrandt and A. Tromba, Mathematics and Optimal Form, Scientific American Books, Inc., New York, 1985.

28. D. Hoffman and H. Karcher, Singly-periodic minimal surfaces derived from crystallographic examples, in preparation.
29. D. Hoffman and W. H. Meeks III, Embedded minimal surfaces of finite topology, *Annals of Mathematics* 131 (1990) 1–34.
30. _____, One-parameter families of embedded complete minimal surfaces of finite topology, preprint.
31. _____, A complete embedded minimal surface with genus one, three ends and finite total curvature, *Journal of Differential Geometry*, 21 (1985) 109–127.
32. _____, A variational approach to the existence of complete embedded minimal surfaces, *Duke Journal of Math.*, 57 (1988) 877–894.
33. _____, The asymptotic behavior of properly embedded minimal surfaces of finite topology, *Journal of the AMS*, 2 (1989) 667–681.
34. D. Hoffman and M. Wohlgemuth, New embedded periodic minimal surfaces of Riemann-type, in preparation.
35. W. Jagy, On Enneper's cyclic minimal surface in higher dimensions, Ph.D. thesis, University of California, Berkeley, 1988.
36. H. Karcher, Embedded minimal surfaces derived from Scherk's examples, *Manuscripta Math.*, 62 (1988) 83–114.
37. _____, Construction of minimal surfaces, *Surveys in Geometry*, 1989, pp. 1–96.
38. _____, The triply periodic minimal surfaces of Alan Schoen and their constant mean curvature companions, *Manuscripta Math.*, 64 (1989) 291–357.
39. N. Korevaar, R. Kusner, and B. Solomon, The structure of complete embedded surfaces with constant mean curvature, *Journal of Differential Geometry*, 30 (1989) 465–503.
40. R. Kusner, Conformal geometry and complete minimal surfaces, *Bulletin of the AMS*, 17 (1987) 291–295.
41. H. B. Lawson, Jr., Lectures on Minimal Submanifolds, Publish or Perish Press, Berkeley, 1971.
42. H. Lewy, On the boundary behavior of minimal surfaces, *Proceedings of the National Academy*, 37 (1951) 103–110.
43. F. J. Lopez and A. Ros, On embedded minimal surfaces of genus zero, preprint, June 1989.
44. W. H. Meeks III and H. Rosenberg, The geometry of periodic minimal surfaces, preprint.
45. _____, The global theory of doubly periodic minimal surfaces, *Inventiones Math.*, 97 (1989) 351–379.
46. W. H. Meeks III and B. White, Minimal surfaces bounded by convex curves in parallel planes, preprint.
47. J. B. Meusnier, Mémoire sur la courbure des surfaces, *Mém. Mathém. Phys. Acad. Sci. Paris, prés. par div. Savans*, 10 (1785) 477–510.
48. F. Morgan, A Beginner's Guide to Geometric Measure Theory, Academic Press, 1987.
49. J. C. C. Nitsche, A characterization of the catenoid, *Journal of Math. Mech.*, 11 (1962) 293–302.
50. _____, On new results in the theory of minimal surfaces, *Bull. Amer. Math. Soc.*, 71 (1965) 195–270.
51. _____, Lectures on Minimal Surfaces, vol. 1, Cambridge University Press, 1989.
52. R. Osserman, Minimal surfaces in the large, *Commentari Mathematici Helvetici*, 35 (1961) 65–76.
53. _____, Global properties of minimal surfaces in E^3 and E^n, *Annals of Math.*, 80 (1964) 340–64.
54. R. Osserman, A Survey of Minimal Surfaces, Dover Publications, New York, 2nd edition, 1986.
55. J. Pitts and H. Rubenstein, Applications of minimax to minimal surfaces and the topology of three-manifolds, in *Proceedings for the Center for Mathematical Analysis*, pages 137–170, Canberra, Australia, 1987.
56. B. Riemann, Ouevres Mathématiques de Riemann, Gauthiers-Villars, Paris, 1898.
57. M. Schiffman, On surfaces of stationary area bounded by two circles, or convex curves, in parallel planes, *Annals of Math*, 63 (1956) 77–90.
58. A. Schoen, Infinite periodic minimal surfaces without self-intersections, Technical Note D-5541, NASA, Cambridge, Mass., May 1970.
59. _____, Uniqueness, symmetry, and embeddedness of minimal surfaces, *Journal of Differential Geometry*, 18 (1983) 791–809.
60. H. A. Schwarz, Gesammelte Mathematische Abhandlungen, volume 1, Springer, Berlin, 1890.
61. L. E. Scriven, Equilibrium bicontinuous structures, *Nature*, 263 (1976) 123–125.
62. S. Smale, An infinite dimensional version of Sard's theorem., *American Journal of Math.*, 87 (1965) 861–866.

63. F. Tomi and A. J. Tromba, Extreme curves bound an embedded minimal surface of disk type, *Math. Z.*, 158 (1978) 137–145.
64. A. J. Tromba, On the number of simply connected minimal surfaces spanning a curve, *Mem. Amer. Math. Soc.* (1977) 194.
65. B. White, New applications of mapping degrees to minimal surface theory, *Journal of Differential Geometry*, 29 (1989) 143–162.
66. _____, Some recent developments in differential geometry, *Mathematical Intelligencer*, 11 (1989) 41–47.

Curvature in the Eighties

ROBERT OSSERMAN, *Stanford University*

ROBERT OSSERMAN wrote a Ph.D. thesis on Riemann surfaces under the direction of Lars V. Ahlfors at Harvard University. He gradually moved from geometric function theory to minimal surfaces, differential geometry, isoperimetric inequalities, and some aspects of partial differential equations and ergodic theory. He has been at Stanford University since 1955 and will be spending half time as Deputy Director of MSRI in Berkeley for a period of three years, starting September 1, 1990.



The notion of curvature is one of the central concepts of differential geometry; one could argue that it is *the* central one, distinguishing the geometrical core of the subject from those aspects that are analytic, algebraic, or topological. In the words of Marcel Berger [**1**, p. 9], curvature is "the N° 1 Riemannian invariant and the most natural. Gauss and then Riemann saw it instantly."

Curvature also plays a key role in physics. The magnitude of a force required to move an object at constant speed along a curved path is, according to Newton's laws, a constant multiple of the curvature of the trajectory. The motion of a body in a gravitational field is determined, according to Einstein, by the curvature of space-time. All sorts of shapes, from soap bubbles to red blood cells, seem to be determined by various curvatures.

Given the vastness of the subject and the limited space available, I have restricted myself to some modest goals. First, to present a few topics that are easily accessible to a reader unfamiliar with the field. Second, to concentrate on results from the past ten years, to demonstrate the current vitality of the subject. And third, to give some feeling for the various forms of curvatures, both intrinsic and extrinsic.

The results we discuss are mainly of two types: inequalities involving integrals of curvatures, and implications of one or another type of curvature being constant. A huge subject that we shall barely touch on is that of consequences of curvature having a given sign or being constrained in other ways.

On the premise that anything published before 1980 can now be considered classic, we shall give explicit references chiefly to recent books and articles, where many further references may be found.

This article is divided into five parts.

1. Some background definitions and terminology.
2. Curvature inequalities for curves and surfaces.
3. Surfaces of constant mean curvature.
4. Higher dimensions.
5. Related results.

The first three sections concentrate on curves and surfaces in euclidean three-space, with a particular focus on recent work of Antonio Ros. A famous theorem of Alexandrov states that "all soap bubbles are spheres". In other words, soap

bubbles, suitably idealized, must have constant mean curvature, and Alexandrov showed that the only compact surfaces of constant mean curvature, embedded in $\mathbb{R}^3$, are spheres. A new approach to the problem was initiated by Ros [1], and a subsequent paper of Montiel and Ros [1] contains a beautiful and elementary proof of Alexandrov's Theorem. It also contains an equally neat proof of a strong generalization to higher dimensions: the constancy of any of a whole series of different curvatures implies, for a hypersurface in $\mathbb{R}^n$, that it must be a hypersphere. A first case, for which partial results had been obtained by Cheng and Yau [1], was settled by Ros in his first paper. That was followed by different proofs of the general theorem by Ros [2] and Korevaar [1]. Montiel and Ros' proof is given in section 4.

Section 5 includes some further results on hypersurfaces of euclidean space and of spheres, as well as some purely Riemannian matters. We describe a recent breakthrough of Szabó [1] on a nearly fifty-year-old conjecture of Lichnerowicz, but do not go into the details of the proof.

I would like to thank particularly Manfredo do Carmo, Jerry Kazdan, Ulrich Pinkall, and Harold Rosenberg, for pointing me in useful directions as I was preparing this paper.

1. Background

For a plane curve one has the familiar definition of curvature:

$$\kappa = \frac{d\alpha}{ds} \tag{1}$$

where α is the angle made by the tangent vector to the curve (in the direction of travel) with a fixed direction, say the positive x-axis, and s measures arc length along the curve.

Note that curvature is assigned not only a magnitude, measuring the rate of deviation from "straight-aheadness," but also a sign, indicating the *direction* of deviation: positive if we veer off to the left, negative to the right.

For a curve in space, we have no way of attributing a sign to curvature. Again using the parameter s of arc length along the curve, and denoting by x the position vector on the curve, we have the unit tangent vector

$$T = \frac{dx}{ds}$$

and the curvature k defined by

$$k = \left|\frac{dT}{ds}\right| = \left|\frac{d^2x}{ds^2}\right|. \tag{2}$$

If the curve happens to lie in a plane, then the two definitions of curvature agree, up to sign:

$$k = |\kappa|. \tag{3}$$

Associated to a surface S in $\mathbb{R}^3$ are a number of curvatures:

$$\text{the } \textit{principal curvatures } k_1, k_2, \tag{4}$$

$$\textit{mean curvature } H = \frac{k_1 + k_2}{2}, \tag{5}$$

$$\textit{Gauss curvature } K = k_1 k_2. \tag{6}$$

As in the case of plane curves, the *signs* of these curvatures play an important role. For all except K, the sign depends on the orientation of the surface S, which in turn is determined locally by a choice of unit normal N. The intersection of S with a normal plane is a plane curve whose curvature, if not zero, is positive when it lies locally on the same side of the tangent line as N, and negative in the opposite case. The principal curvatures k_1, k_2 are the maximum and minimum of the curvatures obtained in this way over all normal planes. Note the following basic facts concerning the curvatures at a point p of S:

$K < 0 \Leftrightarrow k_1$ and k_2 have opposite signs

$\Rightarrow S$ crosses its tangent plane in any neighborhood of p, (a *saddle* point); (7)

$K = 0 \Leftrightarrow$ either (or both) k_1, k_2 is zero (8)

$K > 0 \Rightarrow S$ lies on one side of its tangent plane at p;

the same side as N if $k_1, k_2 > 0$, the opposite side if $k_1, k_2 < 0$. (9)

An important characterization of k_1, k_2, K, comes from the "Gauss map": the map

$$g: S \to \Sigma$$

where Σ is the unit sphere, defined by

$$g: p \mapsto N(p). \tag{10}$$

Then

$$|K| = |\det(dg_p)| = \lim_{U \downarrow \{p\}} \frac{\text{Area } g(U)}{\text{Area } U}, \tag{11}$$

where U is a neighborhood of p and the limit is taken as the neighborhood contracts down to the point p. If $K \neq 0$, then the sign of K is positive if orientation is preserved, and negative if orientation is reversed.

This approach to curvature is the one taken by Gauss in his fundamental paper on surfaces. (See Dombrowski [1].) The names "Gauss map" and "Gauss curvature" are not really justified, however, since Rodrigues [1] not only introduced and studied both of them considerably before Gauss, but made one very important observation that escaped Gauss' notice: the differential of the "Gauss map" is diagonalized by the principal curvature directions. Specifically, if $x(t)$ defines a curve C on the surface and $N(t)$ is the unit normal along C, then $N'(t) = \lambda x'(t)$ for some real λ if and only if the plane spanned by $x'(t)$ and $N(t)$ intersects S in a curve whose curvature is k_1 or k_2. The corresponding tangent direction $x'(t)$ is then called a *principal curvature direction*. Said differently, if u_1, u_2 are local parameters near a point p of S such that at p, they are in the principal curvature directions, then

$$\begin{aligned} \frac{\partial N}{\partial u_1} &= -k_1 \frac{\partial x}{\partial u_1} \\ \frac{\partial N}{\partial u_2} &= -k_2 \frac{\partial x}{\partial u_2}. \end{aligned} \tag{12}$$

Note that an immediate consequence is that

$$\frac{\partial N}{\partial u_1} \times \frac{\partial N}{\partial u_2} = k_1 k_2 \frac{\partial x}{\partial u_1} \times \frac{\partial x}{\partial u_2}. \tag{13}$$

Since the area of the portion of S corresponding to a domain D in the u_1, u_2-plane is given by

$$A = \int_D \left| \frac{\partial x}{\partial u_1} \times \frac{\partial x}{\partial u_2} \right| du_1\, du_2 \tag{14}$$

and the corresponding area of the image on the unit sphere is

$$\hat{A} = \int_D \left| \frac{\partial N}{\partial u_1} \times \frac{\partial N}{\partial u_2} \right| du_1\, du_2, \tag{15}$$

it follows from (11) and (13) that $|K| = |k_1 k_2|$. It also follows from (13) that the orientation is preserved or reversed according as $k_1 k_2$ is positive or negative. Thus the expression (6) for K as $k_1 k_2$ is equivalent to the description following (11) in terms of the Gauss map.

A final note regarding orientation. Given any local parameters u_1, u_2 on a surface in S, one has an induced orientation on the surface from the standard orientation in the $u_1 u_2$-plane. That induced orientation on S corresponds to the choice of unit normal:

$$N = \frac{\dfrac{\partial x}{\partial u_1} \times \dfrac{\partial x}{\partial u_2}}{\left| \dfrac{\partial x}{\partial u_1} \times \dfrac{\partial x}{\partial u_2} \right|}. \tag{16}$$

These facts and formulas are all that is needed in order to follow the arguments in sections 2 and 3 below. Section 4 involves some generalizations to higher dimensions.

First, if S is a hypersurface in $\mathbb{R}^n$, p a point of S, N a unit normal at p, then for any curve $x(s)$ on S with $x(0) = 0$, if s is the arclength parameter, then the curvature vector $x''(0)$ has a component in the direction N that depends only on the unit tangent vector $T = x'(0)$. In particular, if the curve C is the intersection of S with the plane through p spanned by N and a given unit tangent vector T, then the quantity obtained in the above fashion is just the ordinary curvature of the plane curve C. As the curve $x(s)$ varies, we obtain a set of values

$$x''(s) \cdot N = k(T) \tag{17}$$

called the *normal curvatures* of S at p. The right-hand side of (17) is the restriction to the unit sphere of a quadratic form Q on the tangent space. It follows that there exists an orthonormal basis $e_1, \ldots e_{n-1}$ of the tangent space diagonalizing Q. The directions $e_1, \ldots, e_{n-1}$ are called the *principal curvature directions* at p, and the corresponding quantities

$$k_1 = k(e_1), \ldots, k_{n-1} = k(e_{n-1}) \tag{18}$$

are the *principal curvatures* of S at p.

Just as in $\mathbb{R}^3$ one can consider the Gauss map

$$g: p \mapsto N(p), \tag{19}$$

whose differential

$$dg_p: x'(t) \mapsto N'(t), \qquad (x(t) = p) \tag{20}$$

again satisfies Rodrigues' equations

$$dg_p(e_i) = -k_i e_i, \qquad i = 1, \ldots, n-1. \tag{21}$$

We have the *mean curvature*

$$H = \frac{1}{n-1}(k_1 + \cdots + k_{n-1}) = -\frac{1}{n-1} tr(dg_p) \tag{22}$$

and the *Gauss-Kronecker curvature*

$$k_1 \cdots k_{n-1} = (-1)^{n-1} \det(dg_p), \tag{23}$$

but we also have the other symmetric functions of the principal curvatures, called the *higher-order mean curvatures*. We denote by H_j the jth-*order mean curvature*, normalized so that

$$\prod_{j=1}^{n-1} (1 + tk_j) = \sum_{j=0}^{n-1} \binom{n-1}{j} H_j t^j. \tag{24}$$

Thus, $H_1 = H$, the mean curvature, and H_{n-1} is the Gauss-Kronecker curvature. For $n = 3$, that is all there is, but in higher dimensions, there are intermediate ones. Among those, H_2 plays a special role. Except for a numerical constant, H_2 equals the *scalar curvature*, an important Riemannian invariant. Before describing it, we note that the remainder of the background material in this section is only needed for certain parts of the discussion in section 5 below.

We now move from the special case of a hypersurface to the general case of an m-dimensional submanifold M of $\mathbb{R}^n$. The important distinction made by Gauss in the case of surfaces in $\mathbb{R}^3$ also holds in all generality. *Intrinsic quantities* of M are those that are determined by measurements along M itself, while *extrinsic* ones depend on the way M lies in $\mathbb{R}^n$. Riemann introduced the notion of a *Riemannian manifold* in which lengths are defined by a formula like the one that holds for a submanifold of $\mathbb{R}^n$. About one hundred years later, John Nash showed that although the idea was new (and extremely fruitful) the objects themselves were not, in that no new manifolds arise that are not already among those that occur as submanifolds of $\mathbb{R}^n$.

For surfaces in $\mathbb{R}^3$ it is clear that the principal curvatures and mean curvature are not intrinsic, since rolling a sheet of paper into a cylinder alters them while leaving distances along the surface unchanged. The big surprise, Gauss' *Theorema Egregium*, was that the *product* $k_1 k_2$, the Gauss curvature, *is* intrinsic.

For an m-dimensional Riemannian manifold M, which we may if we like, by Nash's embedding theorem, consider as lying in some $\mathbb{R}^n$, Riemann defined at each point p the *sectional curvatures* as follows. Pick any pair of independent tangent vectors to M at p, say v and w. For every unit vector $u = \lambda v + \mu w$, there is a unique geodesic in M starting at p, with tangent vector u. The set of all such geodesics, as u describes the unit circle in the plane spanned by v and w, sweep

out a surface whose Gauss curvature at p is called the *sectional curvature*

$$K_\Pi = K[v, w] \tag{25}$$

of the plane Π spanned by v and w.

Let $e_1, \ldots, e_m$ be an orthonormal basis of the tangent space to M at p. Then the quantity

$$\tau = 2 \sum_{1 \le i < j \le m} K[e_i, e_j] \tag{26}$$

is independent of the choice of basis, and is called the *scalar curvature* of M at p.

Returning to the special case of a hypersurface in $\mathbb{R}^n$, we may choose the e_i to be principal curvature directions at p. Then a computation very similar to the one that Gauss used to prove that K is intrinsic leads to the equation

$$K[e_i, e_j] = k_i k_j, \tag{27}$$

so that the scalar curvature τ may be expressed in terms of the principal curvatures k_i by

$$\tau = 2 \sum_{1 \le i < j \le n-1} k_i k_j, \tag{28}$$

or by (24):

$$\tau = (n-1)(n-2) H_2. \tag{29}$$

With that as background, we can proceed to the results.

2. Curvature inequalities; curves and surfaces

We start with a classical theorem of Fenchel for a curve in $\mathbb{R}^3$ with curvature k.

THEOREM 2.1. *For a simple closed curve C in $\mathbb{R}^3$,*

$$\int_C k \, ds \ge 2\pi, \tag{30}$$

with equality if and only if C is a convex plane curve.

We give a slightly modified version of the proof in the book by do Carmo [1]. It has the unusual feature of "reducing" a theorem on curves to one on surfaces.

Let

$$K^+ = \max\{K, 0\}.$$

In other words, K^+ is the function on a surface whose value equals the Gauss curvature K wherever K is positive and is zero wherever K is negative or zero.

LEMMA 2.2. *For a compact surface S in $\mathbb{R}^3$,*

$$\int_S K^+ \, dA \ge 4\pi. \tag{31}$$

Proof. The left-hand side of (31) represents the area of the image under the Gauss map (counting multiplicities) of the part of S where $K \ge 0$. Hence, it is sufficient to show that the image covers the whole unit sphere. For that, it suffices to take your surface, as Heinz Hopf used to say, and "set it down on the floor." The first point of S where it touches the floor has a normal vector pointing

vertically downward (assuming that we are using the orientation of the surface corresponding to the outer normal) and has Gauss curvature $K \geq 0$, because at a point where $K < 0$, the surface is saddle-shaped and crosses its tangent plane, by (7). But the "floor" could be in any direction, and thus we get a point with $K \geq 0$ corresponding to every normal direction. Since the area of the unit sphere is 4π, the lemma is proved.

To prove Fenchel's theorem, we apply the lemma to a thin tube around the given curve. In fact the proof is simply a computation showing that for such a tube, the left-hand side of (31) is equal to twice the left-hand side of (30) over the original curve.

Let the curve C be given by

$$x(s), \qquad 0 \leq s \leq L,$$

where s is the parameter of arc length. At each point $x(s)$ of C, we set

$$e_1(s) = x'(s),$$

the unit tangent vector. Let $e_2(s)$ be a smooth unit vector field such that

$$e_2(s) \cdot e_1(s) \equiv 0$$

and set

$$e_3(s) = e_1(s) \times e_2(s).$$

Thus, e_2, e_3 span the normal space to the curve, and the tube of radius r around C is the surface

$$y(s,t) = x(s) + rN(s,t),$$

where

$$N(s,t) = \cos te_2(s) + \sin te_3(s).$$

Setting

$$e_i'(s) = \sum_{j=1}^{3} a_{ij}(s)e_j(s), \qquad i = 1,2,3,$$

and using $a_{ij} = -a_{ji}$ (by differentiating the equation $e_i(s) \cdot e_j(s) \equiv 0$), we find that

$$\frac{\partial y}{\partial t} \times \frac{\partial y}{\partial s} = r[1 + r(a_{21}\cos t + a_{31}\sin t)]N.$$

Thus, for small positive r, the surface $y(s,t)$ is a regular surface, and its unit normal is precisely the vector $N(s,t)$. Furthermore,

$$\frac{\partial N}{\partial t} \times \frac{\partial N}{\partial s} = -(a_{21}\cos t + a_{31}\sin t)N.$$

Since the definition of K amounts to the equation

$$\frac{\partial N}{\partial t} \times \frac{\partial N}{\partial s} = K\frac{\partial y}{\partial t} \times \frac{\partial y}{\partial s},$$

it follows that

$$K = -\frac{a_{21}\cos t + a_{31}\sin t}{r[1 + r(a_{21}\cos t + a_{31}\sin t)]}.$$

But the curvature k of C is

$$k = |x''(s)| = |e_1'(s)| = |a_{12}e_2 + a_{13}e_3| = |a_{21}e_2 + a_{31}e_3| = \sqrt{a_{21}^2 + a_{31}^2}.$$

Thus $k(s) = 0$ implies $K(s,t) \equiv 0$ for $0 \le t \le 2\pi$. At each point where $k(s) > 0$, we may set $a_{21} = k\cos\theta$, $a_{31} = k\sin\theta$, so that

$$K = -\frac{k\cos(t-\theta)}{r[1 + rk\cos(t-\theta)]}.$$

Thus $K > 0$ for $\pi/2 < t - \theta < 3\pi/2$. Since

$$\left|\frac{\partial y}{\partial t} \times \frac{\partial y}{\partial s}\right| = r[1 + rk\cos(t-\theta)],$$

for any interval $a < s < b$ where $k(s) > 0$, we have

$$\int K^+ \, dA = \int_a^b \left[\int_{\theta+\pi/2}^{\theta+3\pi/2} [-k(s)\cos(t-\theta)]\, dt\right] ds = 2\int_a^b k(s)\, ds,$$

and for the entire tube, we get

$$\int K^+ \, dA = 2\int_0^L k(s)\, ds.$$

Fenchel's inequality (30) now follows immediately from (31). A closer analysis shows that equality holds only in the case indicated. (For details, see do Carmo [**1**], p. 400.)

We next ask how Fenchel's inequality extends to surfaces. The answer depends on which curvature we adopt in place of k.

THEOREM 2.3. *Let S be a compact surface embedded in $\mathbb{R}^3$. Then*

$$\int_S |K| \, dA \ge 4\pi, \tag{32}$$

with equality if and only if S bounds a convex domain in $\mathbb{R}^3$.

Proof. Using Lemma 2.2,

$$\int_S |K| \, dA \ge \int_S K^+ \, dA \ge 4\pi.$$

If $K < 0$ at any point, then $K < 0$ in a neighborhood, and the first inequality is strict. Hence equality implies $K \ge 0$ everywhere and by a theorem proved first by Hadamard when $K > 0$, and later by Chern and Lashof [**1**] in general, S must bound a convex domain.

If, instead of K, we consider the mean curvature H, we find several possible formulations.

THEOREM 2.4 (Willmore). *Let S be a compact surface in $\mathbb{R}^3$. Then*

$$\int_S H^2 \, dA \ge 4\pi \tag{33}$$

with equality if and only if S is a standard sphere.

Proof.

$$H^2 = \left(\frac{k_1 + k_2}{2}\right)^2 = k_1 k_2 + \left(\frac{k_1 - k_2}{2}\right)^2 \geq K, \tag{34}$$

so that

$$\int_S H^2\, dA \geq \int_S K^+\, dA \geq 4\pi$$

by Lemma 3.2. Furthermore, inequality is strict unless $H^2 \equiv K$, in which case $k_1 \equiv k_2$ so that every point of S is umbilic. It is then a classical result (for example, do Carmo [**1**], p. 147) that S is a standard sphere.

A question that has received a great deal of attention in the past decade is the *Willmore Conjecture*. If S is topologically a torus, then

$$\int_S H^2\, dA \geq 2\pi^2. \tag{35}$$

This conjecture has been verified for many cases, including when S is a tube about some curve, or of certain conformal type. There are also higher-dimensional versions, chiefly due to B.-Y. Chen. For these and other results on the conjecture, see Bryant [**1**], and the books by Willmore [**1**] and Chen [**1**]. More recently, Langer and Singer [**1**] proved the conjecture for tori of revolution. Leon Simon [**1**] obtained the important result that there exists an immersed torus minimizing the left side of (35), and a result of Li and Yau [**1**] guarantees that such a surface must be embedded. Ferus, Pedit, Pinkall, and Sterling [**1**] provide, in some sense, a complete classification of Willmore tori; that is, stationary surfaces for the integral in (35).

An inequality of a somewhat different character is one obtained by Minkowski for a surface S bounding a convex domain D of volume V. We may write the inequality as

$$\int_S H\, dA \leq \frac{A^2}{3V}, \tag{36}$$

where A is the area of S. Equality holds (for smooth surfaces) only when S is a standard sphere.

We come finally to the recent results of Ros. He proved a new inequality, in a way dual to (36).

THEOREM 2.5 (Ros [**1**]). *Let S be a compact embedded surface in $\mathbb{R}^3$ bounding a domain D of volume V. If the mean curvature H of S is positive everywhere, then*

$$\int_S \frac{1}{H}\, dA \geq 3V. \tag{37}$$

Equality holds if and only if S is a standard sphere.

Ros' original proof uses results of Reilly [**1**]. A subsequent paper of Montiel and Ros [**1**], based on ideas of Heintze and Karcher [**1**], uses a more geometric approach.

Proof. The idea is to use a well-known formula (whose derivation we include at the end of the proof) for the volume of a one-sided shell on the inside of S, where we allow for a shell of variable thickness (somewhat like an oyster shell). In other words, we compute the volume of the domain obtained by starting at each point p of S and going a distance $h(p)$ along the inner normal to S. The volume of this shell is given by

$$V = \int_S F\,dA \tag{38}$$

where

$$F = \int_0^{h(p)} |(1 - k_1 t)(1 - k_2 t)|\,dt = \int_0^{h(p)} |1 - 2Ht + Kt^2|\,dt, \tag{39}$$

provided there is not overlap; that is, provided no portion of the shell is counted twice along normals from different parts of S. To prevent overlap we define $h(p)$ as follows:

$$h(p) = \sup\{r : \text{the point } p \text{ is the unique nearest point on } S \text{ to the point } q \text{ at distance } r \text{ from } p \text{ along the normal to } S \text{ at } p\}. \tag{40}$$

From the definition of $h(p)$ it follows that the interior of the shell of variable width $h(p)$ has no overlap; every point of the interior lies on a unique normal to S. On the other hand, every point of D lies in the closed shell. Namely, given q in D, let d be the distance from q to S. Then the open ball $B_q(d)$ of radius d centered at q contains no point of S, but there is at least one point p of S on its boundary. For any point q' on the radius from q to p, if r is the distance from q' to p then the closed ball $\overline{B_{q'}(r)}$ lies in the interior of $B_q(d)$ except for the point p. It follows that p is the unique point of S realizing the distance r from q' to S. By the definition (40) of $h(p)$, we have $d \leq h(p)$. Hence q lies in the closed shell, as claimed.

We conclude that the volume V of D is precisely equal to the volume of the shell given by (38), since all points of D are covered and the only points covered twice are in the image of the boundary, which is lower dimensional and contributes nothing to the volume.

The key to the argument is now the elementary but crucial observation:

$$\frac{1}{h(p)} \geq \max\{k_1(p), k_2(p)\} \tag{41}$$

for every point p on S. Namely, let q be the point at distance $h(p)$ from p along the normal. Then the open ball $B_q(h(p))$ cannot contain any points of S, since if there were such a point p', then its distance to q would be less than $h(p)$; for all points q' on the radius from q to p, sufficiently near q, the distance from q' to p' would then be less than the distance from q' to p, contradicting the definition of $h(p)$. But since the surface S lies outside the sphere $S_q(h(p))$, the surface and the sphere are tangent at p, and the normal curvatures of S at p are all bounded above by the normal curvatures of the sphere, which equal $1/h(p)$. This proves (41).

It follows from (41) that each of factors

$$(1 - k_1 t), \quad (1 - k_2 t)$$

in (39) is non-negative for $0 \le t \le h(p)$, so that

$$\int_0^{h(p)} |(1 - k_1 t)(1 - k_2 t)|\, dt = \int_0^{h(p)} (1 - k_1 t)(1 - k_2 t)\, dt. \tag{42}$$

By the inequality of the geometric and arithmetic mean,

$$(1 - k_1 t)(1 - k_2 t) \le (1 - Ht)^2 \tag{43}$$

and by (41)

$$\frac{1}{h(p)} \ge H(p)$$

so that from (39)

$$F \le \int_0^{1/H} (1 - Ht)^2\, dt = \frac{1}{3H}$$

and by (38)

$$V \le \int_S \frac{1}{3H}\, dA,$$

proving (37). For equality to hold, it must hold in (43), which means that $k_1 = k_2$. But then by the classical result referred to earlier in the proof of Theorem 2.4, S must be a sphere. This proves the theorem.

For completeness, we show how one derives equations (38) and (39). It is sufficient to carry out the computation locally on S, since one can then add up to obtain the integral over the whole surface. So suppose we have S given locally in parametric form:

$$S\colon x(u_1, u_2),\ (u_1, u_2) \in D \subset \mathbb{R}^2.$$

At each point we have the unit normal $N(u_1, u_2)$, and we parametrize our "one-sided shell" by

$$y(u_1, u_2, u_3) = x(u_1, u_2) + u_3 N(u_1, u_2), \qquad (u_1, u_2) \in D,\ 0 < u_3 < h(u_1, u_2).$$

The volume of this shell is then given by

$$\int\int_D \left[\int_0^{h(u_1, u_2)} \left| \frac{\partial(y_1, y_2, y_3)}{\partial(u_1, u_2, u_3)} \right| du_3 \right] du_1\, du_2,$$

provided the correspondence $(u_1, u_2, u_3) \mapsto (y_1, y_2, y_3)$ is one-to-one (which the definition (40) of $h(p)$ is designed to guarantee). To evaluate the integrand, we note that if u_1, u_2 are principal curvature directions, then Rodrigues' equations (12) imply

$$\frac{\partial y}{\partial u_1} = \frac{\partial x}{\partial u_1} + u_3 \frac{\partial N}{\partial u_1} = (1 - k_1 u_3) \frac{\partial x}{\partial u_1},$$

$$\frac{\partial y}{\partial u_2} = (1 - k_2 u_3) \frac{\partial x}{\partial u_2}, \qquad \frac{\partial y}{\partial u_3} = N;$$

$$\frac{\partial y}{\partial u_1} \times \frac{\partial y}{\partial u_2} = (1 - k_1 u_3)(1 - k_2 u_3) \frac{\partial x}{\partial u_1} \times \frac{\partial x}{\partial u_2}$$

$$= \left(1 - 2Hu_3 + Ku_3^2\right) \frac{\partial x}{\partial u_1} \times \frac{\partial x}{\partial u_2},$$

where $-H$ and K are the trace and determinant of the differential of the Gauss map g given by (10); that is, if

$$dg\colon \frac{\partial x}{\partial u_i} \mapsto \frac{\partial N}{\partial u_i} = \sum c_{ij} \frac{\partial x}{\partial u_j},$$

then the trace and determinant of (c_{ij}) is independent of choice of parameters. The equation above for $(\partial y/\partial u_1) \times (\partial y/\partial u_2)$ is, therefore, valid for any parameters u_1, u_2. Then

$$\left|\frac{\partial(y_1, y_2, y_3)}{\partial(x_1, x_2, x_3)}\right| = \left|\left(\frac{\partial y}{\partial u_1} \times \frac{\partial y}{\partial u_2}\right) \cdot \frac{\partial y}{\partial u_3}\right| = |1 - 2Hu_3 + Ku_3^2| \left|\frac{\partial x}{\partial u_1} \times \frac{\partial x}{\partial u_2}\right|,$$

so that from (14)

$$V = \int\int_D \left[\int_0^{h(u_1, u_2)} |1 - 2Hu_3 + Ku_3^2|\, du_3\right] \left|\frac{\partial x}{\partial u_1} \times \frac{\partial x}{\partial u_2}\right| du_1\, du_2$$

$$= \int\int_S \left[\int_0^{h(u_1, u_2)} |1 - 2Hu_3 + Ku_3^2|\, du_3\right] dA,$$

which gives equations (38), (39).

We have confined ourselves in this section to the case of compact surfaces. However there are many interesting results involving integrals of various curvatures over complete, non-compact surfaces. We note in particular a recent paper by Brian White [**1**] and earlier ones by Osserman [**1**], [**2**].

3. Surfaces of constant mean curvature

One consequence of Ros' inequality (37) is a remarkably elementary proof of a famous theorem of Alexandrov: *the only compact embedded surfaces of constant mean curvature in* $\mathbb{R}^3$ *are the standard spheres.*

To see that, Ros makes use of a general equation, valid for any oriented compact immersed surface S in $\mathbb{R}^3$:

$$A = -\int_S HN \cdot x\, dA, \tag{44}$$

where A is the area of S, H its mean curvature, N the unit normal and x the position vector. Equation (44) can be proved in various ways, the simplest, perhaps, being a special case of the first variation formula; if v is any smooth vector field along S, let S_r be the surface defined by

$$y = x + rv \tag{45}$$

and let $A(r)$ be the area of S_r. Then an elementary computation gives

$$A'(0) = -2\int_S HN \cdot v\, dA. \tag{46}$$

In this case, we choose $v = x$, so that (45) reduces to

$$y = (1 + r)x,$$

a uniform stretching of S by the factor $(1 + r)$, so that

$$A(r) = (1 + r)^2 A$$

and

$$A'(0) = 2A. \tag{47}$$

Combining (46) with (47) gives equation (44).

Now restrict to the case where S is an embedded surface of constant mean curvature H bounding a domain D. If we choose N to be the interior normal at each point, then $H > 0$, since S lies inside some large sphere, and contracting the sphere until it makes a first point of contact p with S, we find that both principal curvatures at p with respect to the inner normal must be positive. Applying the divergence theorem we deduce from (44):

$$A = -H\int_S N \cdot x\, dA = H\int_D \operatorname{div} x\, dV = 3HV \tag{48}$$

since

$$\operatorname{div} x \equiv 3.$$

Hence,

$$H = \frac{A}{3V}, \tag{49}$$

which means that equality holds in (37), and by Ros' theorem, S must be a standard sphere.

For a very pretty elementary proof of Alexandrov's theorem, of a more analytic nature, see Reilly [2]. Incidentally, Harold Rosenberg also noticed that equation (49) holds for constant mean curvature surfaces, and pointed out that as a result, equality holds in Minkowski's inequality (36). Since it is known that at least for smooth convex surfaces, equality can hold in (36) only for the sphere, this gives another proof of the original theorem of Liebmann, the precursor of Alexandrov's, that a *convex* surface of constant mean curvature must be a sphere, and could conceivably yield the full Alexandrov theorem if a suitable extension of Minkowski was found.

A generalization of Liebmann's theorem in a different direction than Alexandrov's was given by Heinz Hopf. He showed that any immersion of a surface, topologically a sphere, with constant mean curvature in $\mathbb{R}^3$ must be a standard sphere. In conjunction with Alexandrov's theorem, that led to one of the most stubborn questions in the field during a period of several decades: must every *immersed* compact surface of constant mean curvature in $\mathbb{R}^3$ be a standard sphere? That question, generally known as "Hopf's conjecture," was finally settled in 1986 by Wente [1] who demonstrated the existence of an immersed torus of constant mean curvature. Wente's work inspired a string of further results on compact tori of constant mean curvature (Abresch [2], Spruck [1], Pinkall-Sterling [1]). Using radically different methods, Kapouleas [2] has just shown that there also exist compact immersed surfaces with constant mean curvature of every genus $g \geq 3$. The existence of such surfaces with genus 2 remains an open question.

We may mention just a few further results of interest on constant mean curvature surfaces proved in the last few years.

In the case of compact surfaces, the first-variation formula (46) applied to a normal variation

$$v = fN$$

for an arbitrary function f, gives

$$A'(0) = -2\int_S fH\,dA.$$

If S bounds a domain of volume V, and the varied surfaces S_t bound domains of volume $V(t)$, then

$$V'(0) = \int_S f\,dA.$$

In particular, if the volume is held fixed, so that

$$V(t) \equiv V,$$

then $V'(0) = 0$, and the problem of minimizing surface area while fixing the enclosed volume leads to the analytic formulation of finding a surface S such that

$$\int_S fH\,dA = 0 \quad \text{whenever } \int_S f\,dA = 0. \tag{50}$$

But a standard argument shows that (50) holds if and only if H is constant. A surface S satisfying (50) is called *stationary* with respect to the given class of variations. It is called *stable* if in addition

$$A''(0) \geq 0.$$

Note that condition (50) makes sense for any immersed surface, as does stability.

THEOREM (Barbosa and do Carmo [1]). *A compact stable surface of constant mean curvature is a standard sphere.*

Thus, Wente's surfaces and their successors must all be unstable.

Do Carmo's theorem was extended by Palmer [1], Silveira [1] and Lopez and Ros [1] to complete surfaces, where the stability assumption applies to every compact subdomain and the surface is assumed to have non-zero constant mean curvature. This result, along with many others on complete surfaces of non-zero constant mean curvature, tends to restrict the possibilities for such surfaces. Other such theorems are those of Klotz-Osserman and Hoffman-Osserman-Schoen [1] stating that a complete surface of non-zero constant mean curvature must be a circular cylinder if, in the first case, its Gauss curvature K satisfies $K \leq 0$, and in the second, its image under the Gauss map lies in a closed hemisphere. The latter result was extended by Lopez and Ros [1] who showed that it is sufficient to assume that for some compact set on the surface, the image of its complement under the Gauss map lies in a closed hemisphere. These results are sharp, in that, as Seaman showed [1], there exist examples whose Gauss map covers an arbitrarily narrow strip surrounding the equator: namely certain of the surfaces of revolution, called *Delaunay surfaces*.

A beautiful very recent result by Korevaar, Kusner and Solomon [1] characterizes the Delaunay surfaces as the only doubly-connected surfaces (besides the circular cylinder) properly embedded in $\mathbb{R}^3$ with non-zero constant mean curvature.

In view of the paucity of examples, it represented a major breakthrough when Kapouleas [1] proved the existence of a whole new species of complete surfaces of non-zero constant mean curvature, including examples of infinite genus, and infinitely many topologically distinct ones of every finite genus. Each annular end converges to a Delaunay surface. That turns out not to be an accident, as was demonstrated shortly after in the paper of Korevaar, Kusner and Solomon [1].

They prove:

THEOREM. *Let S be a properly embedded surface of finite topological type with constant non-zero mean curvature. Then each end is asymptotic to a Delaunay surface.*

We have restricted our attention in this section to surfaces of *non-zero* constant mean curvature. Surfaces with $H \equiv 0$, *minimal surfaces*, constitute a whole world of their own, about which one can learn in a number of recent books: Barbosa and Colares [**1**], Bourguignon, Lawson, and Margerin [**1**], Nitsche [**1**], Osserman [**3**], [**4**], and Yang [**1**] in particular.

4. Higher dimensions

Some of the results described in sections 2 and 3, such as Hopf's and Kapouleas' Theorems, use the two-dimensionality of the surfaces studied in an essential way. In fact, Hopf's theorem turns out to be false in general, as was proved by Wu-Yi Hsiang [**1**], who constructed immersed spheres of constant mean curvature in $\mathbb{R}^n$, $n \geq 4$, different from the standard sphere. However, other results extend to hypersurfaces in higher dimensions, in many cases in a straightforward manner. In particular, Ros' fundamental inequality, along with its proof, extends to compact hypersurfaces S in $\mathbb{R}^n$ with $H > 0$ in the form

$$\int_S \frac{1}{H}\, dA \geq nV, \tag{51}$$

where S is assumed to be the boundary of a domain D of volume V, and again, equality holds only for the standard sphere. But also the first variation formula (46) and its consequences, (44) and (48), extend easily to give equality in (51) when H is constant, and hence proves Alexandrov's theorem in $\mathbb{R}^n$.

Ros' main goal, however, was to prove the conjecture that constant *scalar* curvature implies that an embedded compact hypersurface must be a sphere. He did that [**1**] using results of Reilly [**1**], together with the fact that, as follows from (29), constant scalar curvature is equivalent to constant second-order mean curvature. That paper led Ros [**2**] and Korevaar [**1**] independently to a proof of the general theorem that the constancy of *any* of the H_j is possible only for a sphere. Korevaar showed how Alexandrov's original argument for H_1 could be carried over, whereas Ros adapted the reasoning in his earlier paper. Finally, Montiel and Ros [**1**] gave the new and elementary proof of Ros' inequality (51) that we described above.

To complete the proof for the higher order mean curvatures H_j, defined as we noted earlier by the equation

$$\prod_{j=1}^{n-1} (1 + tk_j) = \sum_{j=0}^{n-1} \binom{n-1}{j} H_j t^j,$$

Ros makes use of the Minkowski equations

$$\int_S (H_j + H_{j+1} N \cdot x)\, dA = 0. \tag{52}$$

Note that equation (44) is just the special case $j = 0$ of (52). But the general equations follow from the special case by computing the mean curvature $H(r)$ of a parallel hypersurface S_r to S, which may be expressed explicitly in terms of the quantities H_j on the original surface, and substituting in (44); setting the coefficient of each power of r equal to zero gives Minkowski's equations (see Montiel-Ros

[1]. Incidentally, these "Minkowski equations" were apparently first stated and proved by Kubota for convex hypersurfaces and then obtained in the general case by Hsiung [1]. What Minkowski did was the case $n = 3$, $j = 1$.)

To conclude the proof, Ros shows that if a particular H_j is positive on all of S, then the same holds for $H_1, \ldots, H_{j-1}$, and the following inequalities hold:

$$H_i^{i-1} \leq H_{i-1}^i, \qquad H_i \leq H^i, \qquad i = 1, \ldots, j. \tag{53}$$

Suppose now that H_j is constant on S. Since there exists some point where all the principal curvatures are positive (using the same argument of contact with an enveloping sphere as in $\mathbb{R}^3$), the constant value of H_j is also positive, and inequalities (53) hold. By (52),

$$\int_S H_{j-1}\, dA = -\int_S H_j N \cdot x\, dA = -H_j \int_S N \cdot x\, dA = H_j n V,$$

where we have used the n-dimensional version of (48) in the last step. Invoking the inequalities (53) then yields

$$nV = \frac{1}{H_j}\int H_{j-1}\, dA \geq \frac{1}{H_j}\int H_j^{(j-1)/j}\, dA \geq \frac{A}{H_j^{1/j}} = \int \frac{1}{H_j^{1/j}}\, dA \geq \int \frac{1}{H}\, dA.$$

But this is just the reverse of Ros' inequality (51). Hence equality holds in (51) and by Ros' Theorem, S is a standard sphere.

5. Related results

We now review briefly a few further topics, again with emphasis on some of those where exciting progress has been made in the eighties.

First, another innocuous-sounding question that flowered into a beautiful theory: what are the surfaces for which not only the mean curvature, but each of the individual principal curvatures is constant? That turns out to be such a strong condition, that even locally, any piece of such a surface must lie on a sphere, a plane, or a right circular cylinder. An interesting, and perhaps unexpected feature, is that it is impossible for the principal curvatures of a surface to be, even locally, distinct non-zero constants. Shortly after these results were obtained by Levi-Civita [1], they were extended to hypersurfaces of arbitrary dimension. Segre [1] proved that if all the principal curvatures are constant, then there could be at most two distinct values, and at most one of them different from zero. As a result, the hypersurface must again lie either on a sphere, a hyperplane, or a generalized cylinder $S^k \times \mathbb{R}^{n-k}$, $k = 1, \ldots, n-1$. The whole subject, which was apparently finished shortly after it began, took on an entirely new life with a paper of Cartan [1] who posed the same question for hypersurfaces in a sphere. Before describing the results, let us recall the definitions.

Let C be a curve lying on a sphere S^n in $\mathbb{R}^{n+1}$. If C is given parametrically as $x(s)$ where s is the parameter of arclength and $x(s_0) = p$, then the *curvature vector* of C at p with respect to S^n is the projection of the Euclidean curvature vector $x''(s_0)$ onto the tangent plane of S^n at p. Note that a curve C has zero curvature vector with respect to S if and only if $x''(s)$ is normal to S^n. If that holds on an interval, then one can show that the corresponding arc of C lies on a great circle.

Now let M be an $(n-1)$-dimensional manifold lying on S^n, and p a point of M. By considering the curvature vectors at p with respect to S^n of curves through

p lying on M, we arrive at a set of *principal curvatures of M at p with respect to S^n* in the same manner that we defined in (18) the principal curvatures of a hypersurface in $\mathbb{R}^n$ by using the standard curvatures of curves on the hypersurface at a given point. It is those principal curvatures that Cartan requires to be constant.

In his original paper and a number of subsequent ones, Cartan shows that this class of surfaces, called *isoparametric hypersurfaces*, is far more interesting in S^n than in $\mathbb{R}^n$. The intersection of the cylinder

$$x_1^2 + \cdots + x_m^2 = r_1^2,$$

with the sphere S^n

$$x_1^2 + \cdots + x_{n+1}^2 = r^2 > r_1^2$$

is an isoparametric hypersurface. Cartan showed that every hypersurface of S^n with two distinct (constant) principal curvatures is one of those. They include a famous surface, the *Clifford torus* in S^3. Cartan also showed that there were examples with three distinct curvatures, and that they could all be represented as tubes over a special set of embeddings in certain S^n, including S^4. He then constructed an example with four distinct curvatures in S^5, and was naturally led to ask whether there exist examples with any prescribed number of distinct curvatures. That question remained unanswered for almost forty years, until the appearance in 1980 and 1981 of two remarkable papers by Münzner [**1**]. Among the results he obtains are

1. the only possibilities for the number of distinct curvatures are 1, 2, 3, 4 and 6;

2. every isoparametric hypersurface of S^n lies on an algebraic surface; if there are g distinct curvatures, then it lies on the intersection of S^n with a variety defined by a homogeneous polynomial of degree g.

An illustration of the second result is the theorem of Cartan mentioned earlier that all examples with two curvatures are given by the intersection of the sphere with a cylinder, defined by a quadratic polynomial.

Following the work of Münzner there has been a great burst of activity in the field. For an excellent discussion and references up to 1985, see Chapter 3 of the book by Cecil and Ryan [**1**]. A recent paper of interest with further references is that of Solomon [**1**].

Cartan also considered isoparametric hypersurfaces in hyperbolic space—that is, a space with constant negative sectional curvature. What Cartan showed [**1**] was that isoparametric surfaces in hyperbolic space were limited in the same manner as in Euclidean space, in that they could have at most two distinct principal curvatures.

We should say a word about how principal curvatures are defined for hypersurfaces in a Riemannian manifold, where the metric is only defined intrinsically. First of all, by Nash's embedding theorem, every such manifold M can be thought of as embedded in some Euclidean space $\mathbb{R}^N$, in which case any curve C in M has at each point its ordinary curvature vector $x''(s)$ in $\mathbb{R}^N$. The *geodesic curvature vector* of C at a point is simply the projection of $x''(s)$ onto the tangent space of M at the point. The critical fact is that the geodesic curvature vector can be expressed intrinsically in terms of the Riemannian metric on M. Consequently, one does not need to consider an embedding, but can define curvatures of curves, principal

curvatures of hypersurfaces, and all the corresponding quantities in any Riemannian manifold.

As an example, once one has the principal curvatures of a hypersurface in hyperbolic space, one can define the mean curvature as their average, and all the higher-order mean curvatures as before. One can also define "spheres" in a Riemannian manifold, or "geodesic spheres," as they are usually called, centered at a point p, with radius r, as the set of points reached by going a distance r along an arbitrary geodesic starting at p. For a compact hypersurface M embedded in hyperbolic space, Korevaar [1] and Montiel-Ros [1] prove that if any of the mean curvatures is constant, then M must be a geodesic sphere. The same is true if M lies in a *hemisphere* of S^n. However, for hypersurfaces in the full sphere, the result is *not* true, since one has all the isoparametric examples, which clearly have all mean curvatures constant, since all principal curvatures are constant. One can narrow down the possibilities if one restricts, for example, the sectional curvature of M. (See Walter [1] and further references given there.)

We conclude our discussion of this subject with two further remarks. First, for an isoparametric hypersurface M in a constant curvature space, the parallel hypersurfaces to M are also isoparametric. Second, the family of parallel isoparametric hypersurfaces may be written as the level sets of a function F with the property that both the Laplacian of F and the magnitude of the gradient of F are functions of F itself. In fact this property characterizes isoparametric hypersurfaces and is the origin of their name.

We now enlarge our viewpoint a bit and ask the following question: *for which Riemannian manifolds do all geodesic spheres have constant mean curvature*? It turns out that manifolds with this property arose in the context of "harmonic spaces." These spaces, or manifolds, may be characterized in a number of ways—for example by the property that the mean value of a harmonic function over a geodesic sphere equals its value at the center, just as in Euclidean space (a theorem of Willmore. See also Ruse, Walker and Willmore [1] and Chapter 6 of Besse [1, in particular, 6.19].)

Among manifolds that satisfy the given condition is the famous class of "two-point homogeneous manifolds." They are defined by the property that for every two pairs of points p_1, q_1 and p_2, q_2, such that the distances $d(p_1, q_1), d(p_2, q_2)$ are equal, there is an isometry of the manifold taking p_1 to p_2 and q_1 to q_2. In particular, if $p_1 \neq p_2$, then for any two points q_1, q_2 of a geodesic sphere centered at p_1, there is an isometry of the manifold fixing the point p_1, hence necessarily taking the sphere to itself, and mapping q_1 to q_2. Thus, the mean curvature of the geodesic sphere is necessarily constant. In 1944, Lichnerowitz conjectured that the two-point homogeneous spaces were the *only* harmonic spaces. Very little progress was made for 45 years. Then in 1989, the conjecture was proved by Szabó [1] for all simply connected compact harmonic spaces. The proof draws on much work that had been done on both harmonic spaces and two-point homogeneous spaces separately, as well as some results linking them.

We mention here a few facts about two-point homogeneous spaces in which curvature plays a major role. We start with the celebrated "pinching theorem" of Toponogov, Berger, and Klingenberg:

THEOREM. *Let M be a compact simply-connected manifold whose sectional curvature K satisfies*

$$0 < \delta \leq K \leq 1.$$

Let d be the diameter of M. Then

i) $d \geq \pi$;

ii) *M is homeomorphic to a sphere if any of the following holds*:

a) $\delta > \frac{1}{4}$,

b) $\delta = \frac{1}{4}$ *and* dim M *is odd*,

c) $\delta = \frac{1}{4}$ *and* $d > \pi$;

iii) *if* dim M *is even*, $\delta = \frac{1}{4}$, *and* $d = \pi$, *then M is a two-point homogeneous space.*

For proofs of this and other pinching theorems see the books of Chavel [1] and Cheeger-Ebin [1], or the excellent recent article by Karcher [1].

Regarding part (iii) of the theorem, it turns out that one can list explicitly all the compact two-point homogeneous spaces; they are the standard spheres and real projective spaces, the complex and quaternionic projective spaces, and the projectivized Cayley plane, all with their standard Riemannian metrics.

This same list arose in Cartan's investigations of a certain class of spaces defined by a condition on curvatures. To explain that condition, let us start again with a manifold M embedded in Euclidean space. Let C be a curve in M, and let v be a vector field along C; that is, at each point $x(t)$ of C, $v(t)$ is a tangent vector to M at $x(t)$. Then v is said to be *parallel* along C if $v'(t)$ is normal to M at $x(t)$ for all t. This notion is the parallel translation of Levi-Civita, and like the closely related notion of geodesic curvature, it has the remarkable property that it can be defined intrinsically, independent of whether or how M is embedded in Euclidean space. Furthermore, given C and a tangent vector v_0 at a point $x(t_0)$ of C, there is a unique parallel field $v(t)$ along C with $v(t_0) = v_0$. Finally, if $v(t), w(t)$ are both parallel along C, then $v(t) \cdot w(t)$ is constant. It follows that if Π_0 is a two-dimensional tangent plane at a point $x(t_0)$ of C, then we can define parallel two-planes along C by taking a pair of orthonormal tangent vectors v_0, w_0 spanning Π_0, letting $v(t), w(t)$ be the parallel fields along C extending v_0, w_0 and defining $\Pi(t)$ to be the plane spanned by $v(t)$ $w(t)$, which are necessarily orthonormal for all t. By the properties we have mentioned, this one-parameter family of planes along C exists, is unique, and is independent of the choice of the original pair v_0, w_0.

We come now to the notion introduced by Cartan.

DEFINITION. A Riemannian manifold is *locally symmetric* if the sectional curvature is constant along every family of parallel planes $\Pi(t)$.

A locally symmetric space has *rank one* if none of its sectional curvatures is zero.

THEOREM. *The two-point homogeneous spaces are precisely the Euclidean spaces $\mathbb{R}^n$, the real projective spaces $\mathbb{R}P^n$, and the complete simply-connected locally symmetric spaces of rank one.*

For more on this subject, see the books of Chavel [1], Helgason [1] and Besse [1]. In this last reference, a Compact Rank One Symmetric Space is referred to acronymically as a CROSS. Thus, the theorems of Szabó and Berger referred to above may be stated: a compact simply-connected manifold which is either harmonic, or is even dimensional with diameter π and sectional curvatures pinched between $\frac{1}{4}$ and 1, must be a CROSS.

We conclude this section with a quick review of questions involving manifolds for which one or more of the intrinsically defined curvatures is constant.

First, complete Riemannian manifolds with constant sectional curvature—the "Clifford-Klein space forms"—were studied in detail by Hopf, and their classification was completed by Wolf [1]. The case of three-dimensional manifolds with constant negative curvature has been intensively studied in recent years, with the major impetus coming from Thurston [1], [2].

The requirement of constant sectional curvature imposes severe restrictions on the possible forms of the manifold. One can ask what restrictions arise from requiring that the *scalar* curvature be constant. The fundamental paper here is one of Yamabe from 1960. The question was whether there are *any* restrictions needed to have a metric of constant scalar curvature. Yamabe presented a proof that for compact manifolds there always exists such a metric conformally equivalent to any given metric on the manifold. However, his purported proof turned out to be incomplete, and after partial results by Trudinger, Aubin [1] proved the theorem for all manifolds of dimension $n \geq 6$ that were not conformally flat. To settle the problem completely, a new global type of argument was needed, and that was provided in 1984 by Schoen [1]. A general survey of the problem may be found in Lee and Parker [1]. Subsequent papers by Schoen [2], [3], [4] contain a number of further results, including an analysis of *all* metrics of constant scalar curvature in a given conformal class.

Thus, in contrast to the case of embedded hypersurfaces where, as we saw, constant scalar curvature was possible only for a sphere, for general compact manifolds there are no topological implications whatever. On the other hand, the *sign* of the scalar curvature does have significance. For a wealth of information on that topic we refer to the book of Lawson and Michelson [1].

The sectional curvature and the scalar curvature are the two most immediate curvature quantities that arise in the study of Riemannian manifolds. However, experience has shown that a third—intermediate—curvature is in many respects the "right one" to consider. To define it, we look more closely at the totality of sectional curvatures at a point p of a Riemannian manifold M.

Let v be a unit tangent vector to M at p. Let $H_v = v^{\perp}$ be the hyperplane in the tangent space orthogonal to v. For $w \in H_v$, set

$$Q_v(0) = 0, \qquad Q_v(w) = |w|^2 K(v, w) \text{ for } w \neq 0 \tag{54}$$

where

$$K(v, w) = \text{sectional curvature of the plane spanned by } v \text{ and } w. \tag{55}$$

Then Q_v is a quadratic form on H_v whose eigenvalues we denote

$$\lambda_1(v) \geq \lambda_2(v) \geq \cdots \geq \lambda_{n-1}(v) \tag{56}$$

where $\dim M = n$. The corresponding eigenvectors

$$e_1, \ldots, e_{n-1}$$

form an orthonormal basis of H_v, and

$$\lambda_i(v) = Q_v(e_i) = K(v, e_i)$$

is a sectional curvature of a plane through v, with λ_1 and λ_{n-1} representing the maximum and minimum of such curvatures.

DEFINITION. The *Ricci curvature* $R(v)$ of M at p in the direction v is

$$R(v) = \sum_{i=1}^{n-1} \lambda_i(v). \tag{57}$$

Thus, the Ricci curvature is the trace of the quadratic form Q_v, and since the trace is independent of basis, we can also write

$$R(v) = \sum_{i=1}^{n-1} K(v, e_i) \tag{58}$$

where $e_1, \ldots, e_{n-1}$ is *any* orthonormal basis of H_v.

Geometrically, it would generally make more sense to consider the *average* rather than the *sum* of the sectional curvatures, and in some places that convention is adopted. However, (58) is the more common usage.

DEFINITION. An *Einstein manifold* is a Riemannian manifold with constant Ricci curvature.

The literature on Einstein manifolds is immense—we refer to the excellent book of Besse [**2**]—and there are many recent results of interest involving the Ricci curvature. Let us mention just one: the celebrated theorem of Hamilton [**1**].

THEOREM. *Let M be a compact manifold,* dim $M = 3$. *If there exists a metric on M with positive Ricci curvature, then there also exists one of constant positive sectional curvature.*

Since, as we mentioned, manifolds with constant sectional curvature are completely classified, Hamilton's Theorem implies that in three dimensions, positive Ricci curvature has powerful implications.

In view of the rich theory that grew out of replacing submanifolds of constant mean curvature by the more restrictive class of manifolds with each principal curvature constant—the isoparametric manifolds—it seems natural to consider a kind of intrinsic analog. We ask about manifolds satisfying the following condition:

$(*)$ *the set of eigenvalues $\lambda_1(v), \ldots, \lambda_{n-1}(v)$ in* (56) *is the same for all p and v.*

Clearly, by (57), such manifolds are Einstein manifolds.

Although the analogy with isoparametric surfaces would be an obvious motivation for introducing condition $(*)$, that was not the original source of the notion. It arose from some results in ergodic theory giving a lower bound for entropy of geodesic flow in terms of curvature quantities (Osserman-Sarnak [**1**]) and from a later conjecture about a possible upper bound for the entropy (see Burns and Katok [**1**], Conjecture 5.4). A discussion of these matters with Christopher Croke led to condition $(*)$.

For rank one locally symmetric spaces the values of the $\lambda_i(v)$ in (56) are known explicitly: there are just two distinct values, with ratio $1/4$, and they are independent of p and v. Thus all such spaces satisfy condition $(*)$. In view of some of the results from ergodic theory, we are led to the following:

CONJECTURE. *The only nonflat manifolds satisfying condition $(*)$ are the rank one locally symmetric spaces.*

More details on the background to this conjecture are given in the paper of Chi [**1**], where it is shown that the conjecture does indeed hold in many cases: in particular for dimension 4, $2k + 1$, or $2(2k + 1)$. A later paper of Chi [**2**] settles several further cases. However, the general conjecture appears quite difficult. Of course it is possible that a counterexample may turn up, in which case the class of manifolds satisfying $(*)$ may turn out to be of separate interest.

For an interesting discussion of related questions and many more references, see the papers of Vanhecke [1], [2].

Conclusion

Of the many topics that could have been included in this discussion but were not, let me name just a few: there are the "almost flat manifolds" of Gromov, the classification of manifolds of non-positive curvature by Ballmann, Brin, Eberlein, Burns and Spatzier, and a whole series of results relating curvature bounds to topology. A few references for these subjects are included in the bibliography. I hope that at least the topics touched on here will convey some notion of the role of curvature in geometry, and a sense of its continuing fascination in current research.

Notes added in proof. Both Yu D. Burago and Joel Spruck have pointed out that for the special case of convex hypersurfaces, Ros' inequality (37) is a consequence of Minkowski's (36):

$$3V\int_S H\,dA \le A^2 = \left(\int_S \sqrt{H}\,\frac{1}{\sqrt{H}}\,dA\right)^2 \le \int_S H\,dA \int_S \frac{1}{H}\,dA.$$

The fact that equality holds in both Minkowski's and Ros' inequalities when H is constant turns out for star-shaped domains to be a special case of the equation $3V\overline{H} = A$ that holds for all such domains, where $\overline{H}$ is the volume average of the function obtained by extending the mean curvature H of the surface to be constant along radii. (See Theorem 9.1, p. 234 of Finn [1].)

An interesting additional reference on surfaces of constant mean curvature is the recent paper of Bobenko [1].

Finally, there are strong parallels between embedded surfaces of constant mean curvature in $\mathbb{R}^3$ and metrics of constant scalar curvature on conformally flat manifolds. Alexandrov's Theorem that the sphere is the only embedded compact surface of constant mean curvature has a counterpart in a uniqueness theorem of Obata, and the initial results of Kapouleas [1] were an offshoot of some of Schoen's work [2] related to the Yamabe problem.

REFERENCES

V. Abresch
1. Lower curvature bounds, Toponogov's Theorem, and bounded topology, I and II, *Ann. Sci. École Norm.* Sup. 18 (1985), 651–670 and 20 (1987), 475–502.
2. Constant mean curvature tori in terms of elliptic functions, *J. reine u. angew. Math.* 374 (1987), 169–192.

S. de Almeida and F. Brito
1. Minimal hypersurfaces of S^4 with constant Gauss-Kronecker curvature, *Math. Z.* 195 (1987), 99–107.
2. The geometry of closed hypersurfaces, *Séminaire de théorie spectrale et géométrie* n°6, Chambéry-Grenoble 1987–1988, 109–116.

T. Aubin
1. Équations différentielles non linéaires et problème de Yamabe concernant la courbure scalaire, *J. Math. Pures et Appl.* 55 (1976) 269–296.

W. Ballmann, M. Brin and P. Eberlein
1. Structure of manifolds of nonpositive curvature I. *Ann. of Math.* (2) 122 (1985), 171–203.

W. Ballmann, M. Brin and R. Spatzier
1. Structure of manifolds of nonpositive curvature II. *Ann. of Math.* (2) 122 (1985), 205–235.

W. Ballmann, M. L. Gromov, and V. Schroeder
1. *Manifolds of Nonpositive Curvature*, Progress in Mathematics 61, Birkhäuser, Boston 1985.

J. L. Barbosa and A. G. Colares
1. *Minimal Surfaces in* $\mathbb{R}^3$, Lecture Notes in Mathematics 1195, Springer 1986.
J. L. Barbosa and M. do Carmo
1. Stability of hypersurfaces with constant mean curvature, *Math. Z.* 185 (1984), 339–353.
M. Berger
1. La géométrie métrique des variétés Riemanniennes, in *Élie Cartan et les Mathématiques d'Aujourd'Hui*, Astérisque, N° Hors Série, Société Mathématique de France 1985, pp. 9–66.
M. Berger and B. Gostiaux
1. *Differential Geometry: Manifolds, Curves, and Surfaces*, Springer 1988.
A. L. Besse
1. *Manifolds all of whose Geodesics are Closed*, Ergebnisse der Mathematik und ihrer Grenzgebiete 93, Springer 1978.
2. *Einstein Manifolds*, Ergebnisse der Mathematik und ihre Grenzgebiete, 3. Folge, Vol. 10, Springer 1987.
A. I. Bobenko
1. All constant mean curvature tori in $\mathbb{R}^3$, S^3, H^3, in terms of theta-functions, (preprint) Reihe Mathematik No. 253 (1990), Technische Universität Berlin.
J. P. Bourguignon, H. B. Lawson, Jr., and C. Margerin
1. *Théorie des Variétés Minimales et Applications* (*Minimal Submanifolds*), Séminaire Palaiseau, Astérisque 154–155, Société Math. de France 1987.
R. Brooks
1. A construction of metrics of negative Ricci curvature, *J. Differential Geometry* 29 (1989), 85–94.
R. Bryant
1. A duality theorem for Willmore surfaces, *J. Differential Geometry* 20 (1984), 23–53.
K. Burns and A. Katok
1. Manifolds with non-positive curvature, *Ergodic Theory and Dynamical Systems* 5 (1985), 307–317.
K. Burns and R. Spatzier
1. Manifolds of nonpositive curvature and their buildings, *Inst. Hautes Études Sci. Pub. Math.* No. 65 (1987), 35–39.
P. Buser
1. *Gromov's Almost Flat Manifolds*, Societé Math. de France, Paris 1981.
E. Cartan
1. Familles de surfaces isoparametriques dans les espaces à courbure constante, *Annali di Mat.* 17 (1938), 177–191.
T. E. Cecil and P. J. Ryan
1. *Tight and taut immersions of manifolds*, Research Notes in Mathematics 107, Pitman Advanced Publishing Program, Boston 1985.
I. Chavel
1. *Riemannian Symmetric Spaces of Rank One*, Marcel Dekker, New York 1972.
J. Cheeger and D. G. Ebin
1. *Comparison Theorems in Riemannian Geometry*, North-Holland Publishing 1975.
B.-Y. Chen
1. *Total Mean Curvature and Submanifolds of Finite Type*, World Scientific, Singapore 1984.
B.-Y. Chen and L. Vanhecke
1. Differential geometry of geodesic spheres, *J. Reine Angew Math.* 325 (1981), 28–67.
S.-Y. Cheng and S.-T. Yau
1. Hypersurfaces with constant scalar curvature, *Math. Ann.* 225 (1977), 195–204.
S.-S. Chern and R. K. Lashof
1. On the total curvature of immersed manifolds, *Michigan Math. J.* 5 (1958), 5–12.
Q.-S. Chi
1. A curvature characterization of certain locally rank one symmetric spaces, *J. Differential Geometry* 28 (1988), 187–202.
2. Quaternionic Kähler manifolds and a curvature characterization of two-point homogeneous spaces, *Illinois J. Math.* (to appear).
3. Twistor theory and Blaschke manifolds, *Ann. Global Anal. Geom.* (to appear).
M. do Carmo
1. *Differential Geometry of Curves and Surfaces*, Prentice-Hall, New Jersey 1976.
P. Dombrowski
1. 150 Years after Gauss' "*disquisitiones generales circa superficies curvas*", *Astérisque* 62 (1979).

D. Ferus, F. Pedit, U. Pinkall, and I. Sterling
1. Minimal tori in S^4 and Willmore tori in $\mathbb{R}^3$, (preprint).
R. Finn
2. *Equilibrium Capillary Surfaces*, Springer, 1986.
S. Gallot, D. Hulin, J. Lafontaine
1. *Riemannian Geometry*, Springer 1987; Chapter III: "Curvature".
S. Goldberg
1. *Curvature and Homology*, Dover 1982.
A. Gray
1. *Tubes*, Addison-Wesley 1989.
M. Gromov
1. Curvature, diameter and Betti numbers, *Comment Math. Helv.* 56 (1981), 179–195.
M. L. Gromov and W. P. Thurston
1. Pinching constants for hyperbolic manifolds, *Invent. Math.* 89 (1987), 1–12.
R. Hamilton
1. Three-manifolds with positive Ricci curvature, *J. Differential Geom.* 17 (1982), 255–306.
E. Heintze and H. Karcher
1. A general comparison theorem with applications to volume estimates for submanifolds, *Ann. Sci. École Norm.* Sup. II (1978), 451–470.
S. Helgason
1. *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, New York 1978.
D. Hoffman, R. Osserman, and R. Schoen
1. On the Gauss map of complete surfaces of constant mean curvature in $\mathbb{R}^3$ and $\mathbb{R}^4$, *Comment. Math. Helvetici* 57 (1982), 519–531.
W.-Y. Hsiang
1. Generalized rotational hypersurfaces of constant mean curvature in the Euclidean spaces I, *J. Differential Geometry* 17 (1982), 337–356.
C.-C. Hsiung
1. Some integral formulas for closed hypersurfaces, *Mathematica Scandinavica* 2 (1954), 286–294.
N. Kapouleas
1. Complete constant mean curvature surfaces in Euclidean three-space, *Ann. of Math.* 131 (1990), 239–330.
2. Compact constant mean curvature surfaces in Euclidean three-space, *J. Differential Geometry* (to appear).
H. Karcher
1. Riemannian comparison constructions, *Global Differential Geometry, MAA Studies in Math., Vol. 27*, S.-S. Chern, editor, MAA (1989), 170–222.
N. J. Korevaar
1. Sphere theorems via Alexandrov for constant Weingarten curvature hypersurfaces—appendix to a note of Ros, *J. Differential Geometry* 27 (1988), 221–223.
N. J. Korevaar, R. Kusner, and B. Solomon
1. The structure of complete embedded surfaces with constant mean curvature, *J. Differential Geometry* 30 (1989), 465–503.
J. Langer and D. A. Singer
1. Curves in the hyperbolic plane and mean curvature of tori in 3-space, *Bull. London Math. Soc.* 16 (1984), 531–534.
H. B. Lawson, Jr. and M.-L. Michelson
1. *Spin Geometry*, Princeton University Press 1989.
J. M. Lee and T. H. Parker
1. The Yamabe problem, *Bull. Amer. Math. Soc.* 17 (1987), 37–91.
T. Levi-Civita
1. Famiglie di superficie isoparametriche nell'ordinario spacio euclideo, *Atti Accad. Naz. Lincei. Rend. Cl. Sci. Fis. Mat. Natur* 26 (1937), 355–362.
P. Li and S. T. Yau
1. A new conformal invariant and its applications to the Willmore conjecture and the first eigenvalue of compact surfaces, *Invent. Math.* 69 (1982), 269–291.
F. J. Lopez and A. Ros
1. Complete minimal surfaces with index one and stable constant mean curvature surfaces, Comment. Math. Helvetici 64 (1989), 34–43.

S. Montiel and A. Ros
1. Compact hypersurfaces: the Alexandrov theorem for higher order mean curvatures (to appear in do Carmo 60th birthday volume).

H. F. Münzner
1. Isoparametrische Hyperflächen in Sphären I, II, *Math. Ann.* 251 (1980), 57–71 and 256 (1981), 215–232.

J. C. C. Nitsche
1. *Introduction to Minimal Surfaces*, Cambridge University Press 1989.

R. Osserman
1. Minimal surfaces, Gausss maps, total curvature, eigenvalue estimates, and stability, *The Chern Symposium 1979*, Springer, New York (1980), 199–227.
2. The total curvature of algebraic surfaces, *Contributions to Analysis and Geometry*, Johns Hopkins University Press, Baltimore (1982), 249–257.
3. *A survey of Minimal Surfaces*, 2nd edition, Dover, New York 1986.
4. Minimal surfaces in $\mathbb{R}^3$, *Global Differential Geometry, MAA Studies in Mathematics, Vol. 27,* S.-S. Chern, editor, MAA (1989), 73–98.

R. Osserman and P. Sarnak
1. A new curvature invariant and entropy of geodesic flow, *Inventiones Math.* 77 (1984), 455–462.

B. Palmer
1. Surfaces of constant mean curvature in space forms, Thesis, Stanford University 1986.

U. Pinkall and I. Sterling
1. On the classification of constant mean curvature tori, *Ann. of Math.* 130 (1989), 407–451.

R. C. Reilly
1. Applications of the Hessian operator in a Riemannian manifold, *Indiana Univ. Math. J.* 26 (1977), 459–472.
2. Mean curvature, the Laplacian, and soap bubbles, *Amer. Math. Monthly* 89 (1982), 180–188, 197–198.

O. Rodrigues
1. Recherches sur la théorie analytique des lignes et des rayons de coubure des surfaces, et sur la transformation d'une classe d'intégrales doubles, qui ont un rapport direct avec les formules de cette théorie, Paris, École Polytechnique Corresp. III. (1814–16) 162–182.

A. Ros
1. Compact hypersurfaces with constant scalar curvature and a congruence theorem, *J. Differential Geom.* 27 (1988), 215–220.
2. Compact hypersurfaces with constant higher order mean curvatures, *Revista Matemática Iberoamericana* 3 (1987), 447–453.

H. S. Ruse, A. G. Walker, and T. J. Willmore
1. *Harmonic Spaces*, Cremonese, Rome 1961.

R. M. Schoen
1. Conformal deformation of a Riemannian metric to constant scalar curvature, *J. Differential Geometry* 20 (1984), 479–495.
2. The existence of weak solutions with prescribed singular behavior for a conformally invariant scalar equation, *Comm. Pure Appl. Math.* 41 (1988) 317–392.
3. Variational theory for the total scalar curvature functional for Riemannian metrics and related topics, *Topics in Calculus of Variations*, Lecture Notes in Math. 1365, Springer 1989, 120–154.
4. On the number of constant scalar curvature metrics in a conformal class, (to appear in do Carmo 60th birthday volume).

W. Seaman
1. Helicoids of constant mean curvature and their Gauss maps, *Pacific J. Math.* 110 (1984) 387–396.

B. Segre
1. Famiglie di ipersuperficie isoparametrische negli spazi euclidei ad un qualunque numero di dimensioni, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur* 27 (1938), 203–207.

K. Shiohama, T. Sakai, and T. Sunada, Eds.
1. *Curvature and Topology of Riemannian Manifolds*, Proceedings of the 17th International Taniguchi Symposium in Katata, Japan, Aug. 26–31, 1985; Springer Lecture Notes 1201, 1986.

A. M. da Silveira
1. Stable complete surfaces with constant mean curvature, *Math. Ann.* 277 (1987), 629–638.

L. Simon
1. Existence of Willmore surfaces, *Miniconference on Geometry and Partial Differential Equations*, Proc. Centre Math. Analysis, Australian National University, Vol. 10, 1985, 187–216.

B. Solomon
1. The harmonic analysis of cubic isoparametric minimal hypersurfaces I and II, *Amer. J. Math.* 112 (1990), 157–241.

J. Spruck
1. The elliptic sinh-Gordon equation and the construction of toroidal soap bubbles, Lecture Notes in Mathematics 1340, Springer 1986, 275–301.

Z. I. Szabó
1. The Lichnerowicz conjecture on harmonic manifolds, *J. Differential Geometry* 31 (1990), 1–28.

W. P. Thurston
1. *The Geometry and Topology of three-manifolds*, Princeton Math. Dept. Notes 1979.
2. Hyperbolic structures on 3-manifolds I: Deformation of acylindrical manifolds, *Ann. of Math.* 124 (1986), 203–246.

L. Vanhecke
1. Geometry and symmetry, *Advances in Differential Geometry and Topology*, World Scientific, Singapore 1990.
2. Geometry in normal and tubular neighborhoods, *Proceedings of the Workshop on Differential Geometry and Topology*, Cala Gonone 1988, Univ. Cagliari (to appear).

R. Walter
1. Compact hypersurfaces with a constant higher mean curvature function, *Math. Ann.* 270 (1985), 125–145.

H. Wente
1. Counter-example to the Hopf conjecture, *Pac. J. Math.* 121 (1986), 193–244.

B. White
1. Complete surfaces of finite total curvature, *J. Differential Geometry* 26 (1987), 315–326; correction: Vol. 28 (1988), 359–360.

T. Willmore
1. *Total Curvature in Riemannian Geometry*, Ellis Horwood, Chichester 1982.

J. A. Wolf
1. *Spaces of Constant Curvature*, 4th Ed., Publish or Perish, Berkieley 1977.

K. Yang
1. *Complete and Compact Minimal Surfaces*, Kluwer Academic, Dordrecht 1989.

Conway's Tiling Groups

WILLIAM P. THURSTON, *Princeton University*

DR. THURSTON'S numerous distinctions and honors include the Oswald Veblen Prize in Geometry, the Alan T. Waterman Award of the NSF, and the Fields Medal. His Ph.D. is from Berkeley, in 1972.



1. Introduction

John Conway discovered a technique using infinite, finitely presented groups that in a number of interesting cases resolves the question of whether a region in the plane can be tessellated by given tiles. The idea is that the tiles can be interpreted as describing relators in a group, in such a way that the plane region can be tiled, only if the group element which describes the boundary of the region is the trivial element 1.

A convenient way to describe the construction is by means of the *Cayley graph or graph* of a group. If G is a group, then its graph $\Gamma(G)$ with respect to generators $g_1, g_2 \ldots, g_n$ is a directed graph whose vertices are the elements of the group. For each vertex $v \in \Gamma(G)$, there will be n outgoing edges, labeled by the generators, and n incoming edges: the edge labeled g_i connects v to vg_i.

It is convenient to make a slight modification of this picture when a generator g_i has order 2. In that case, instead of drawing an arrow from v to vg_i and another arrow from vg_i back to v, we draw a single undirected edge labeled g_i. Thus, in a drawing of the graph of a group, if there are any undirected edges, it is understood that the corresponding generator has order 2.

The graph of a group is automatically homogeneous: for every element $g \in G$, the transformation $v \to gv$ is an automorphism of the graph. Every automorphism of the labeled graph has this form. This property characterizes graphs of groups: a graph whose edges are labeled by a finite set F such that there is exactly one incoming and one outgoing edge with each label at each vertex is the graph of a group if and only if it admits an automorphism taking any vertex to any other.

Whenever R is a relator for the group, that is, a word in the generators which represents 1, then if you start from $v \in \Gamma$ and trace out R, you get back to v again. If G has presentation

$$G = \langle g_1, g_2, \ldots, g_n | R_1 = 1, R_2 = 1, \ldots, R_k = 1 \rangle,$$

the graph $\Gamma(G)$ extends to a 2-complex $\Gamma^2(G)$: sew k disks at each vertex of $v \in \Gamma(g)$, one for each relator R_i, so that its boundary traces out the word R_i. An exception is made here for relations of the form $g_i^2 = 1$, since this relation is already incorporated by drawing g_i as an undirected edge. The 2-complex $\Gamma^2(G)$ is simply-connected: that is, every loop in $\Gamma^2(G)$ can be contracted to a point. In fact, if the loop is an edge path, the sequence of edges it follows describes a word in the

generators. The fact that the path returns to its starting point means that the word represents the identity. A proof that this word represents the identity by making substitutions using the relations R_i can be translated geometrically into a homotopy of the path in $\Gamma^2(G)$.

As a very simple example, the symmetric group S_3 is generated by the transpositions $a = (12)$ and $b = (23)$. They satisfy the relation $(ab)^3 = 1$. The graph is a hexagon, with undirected edges, alternately labeled a and b.

A slightly more complicated example is S_4. It is generated by three elements $a = (12)$, $b = (23)$, and $c = (34)$. A presentation is

$$S_4 = \langle a, b, c | a^2 = b^2 = c^2 = 1, (ab)^3 = (bc)^3 = (ac)^2 = 1 \rangle.$$

To construct its graph, first make some copies of the ab hexagon for the S_3 subgroup generated by a and b, and similarly make some copies of bc hexagons. The subgroup generated by a and c is $\mathbf{Z}_2 \times \mathbf{Z}_2$, and its graph is a square with edges labeled alternately a and c. Make copies also of ac-squares. Take one copy of each polygon, and fit them together around a vertex, gluing an a edge to an a edge, *etc.* Around the perimeter of this figure, keep gluing on a copy of the polygon that fits. If you do this systematically, layer by layer, you will have constructed a polyhedron—it is a truncated octahedron. All the edges from the underlying octahedron are labeled b, while the squares produced by truncating the vertices are labeled $acac$.

The reader may enjoy working out the graph of the alternating group A_5, using generators $a = (12)(34)$, and $b = (12345)$. Note that they satisfy the relations $b^5 = 1$ and $ab^3 = (135)^3 = 1$. Try kicking around the construction, with white $ababab$ hexagons and black $bbbbb$ pentagons.

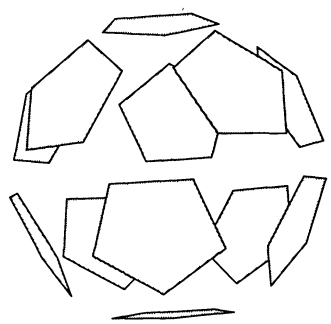


FIG. 1.1. Soccerball. A soccerball is constructed from 12 pentagons, obtained by rotating and shrinking the faces of a regular dodecahedron, together with 20 hexagons centered at the vertices of the dodecahedron.

Of course, graphs of groups don't always work out so nicely or so easily, but often, for simple presentations, they can be worked out, and they tend to have a nice geometric flavor.

2. Lozenges

We will begin with a relatively easy tiling problem. Suppose we have a plane ruled into equilateral triangles, and a certain region R bounded by a polygon π

whose edges are edges of the equilateral triangle network. When can R be tiled by figures, let us call them lozenges, formed from two adjacent equilateral triangles?

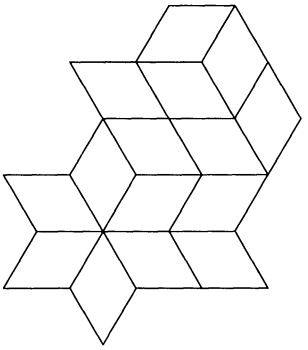


FIG. 2.1. A region tiled by lozenges. A portion of an equilateral triangle subdivision of the plane, tiled by lozenges.

To analyze this problem, we first establish a labeling convention. We arrange the triangulation of the plane so that one set of edges is parallel to the x-axis, or at 0°. Label these directed edges a, label b the directed edges pointing at 120°, and c the edges pointing at 240°. This labeling is homogeneous, so it is the graph of a group A. We can read off relators for A by tracing out the boundary curves of triangles: A satisfies $abc = 1$ and $cba = 1$. If desired, the first relation could be used to eliminate c; the second relation then says that $ba = ab$. The group A is $\mathbf{Z} + \mathbf{Z}$, as we could have seen anyway by its action on the plane.

The shape of the polygon π is determined by the sequence of edges it traces out; this is a word in the generators a, b, c of A. Rather than thinking of it as a word, we prefer to think of it as an element $\alpha(\pi)$ in the free group F with generators a, b, c. The fact that π closes up is equivalent to the condition that the homomorphism $F \to A$ send $\alpha(\pi)$ to the identity.

If a lozenge is placed in the triangular network, its boundary can be traced by one of three elements, depending on its orientation: that element is either $L_1 = aba^{-1}b^{-1}$, $L_2 = bcb^{-1}c^{-1}$, or $L_3 = cac^{-1}a^{-1}$. The precise word depends on the starting point on the boundary of the lozenge, but starting from a different vertex only changes the word by a circular permutation; the two choices give conjugate elements of F. The *lozenge group* L is defined by these relators, that is

$$L = \langle a, b, c | L_1 = L_2 = L_3 = 1 \rangle.$$

Actually, the three relations say that the three generators commute with each other, so that $L = \mathbf{Z}^3$.

We claim that if the region R can be tiled by lozenges, then the image $I(\pi)$ of $\alpha(\pi)$ in L must be trivial. In fact, suppose that we have such a tiling. If R consists of a single tile, the claim is immediate. Otherwise, find a simple arc in R which cuts R into two tiled subregions R_1 and R_2. By induction, we may assume that $I(\pi_1)$ and $I(\pi_2)$ are both trivial, where π_i is a polygonal curve tracing around ∂R_i. But $I(\pi) = I(\pi_1) * I(\pi_2)$, so $I(\pi)$ is also trivial.

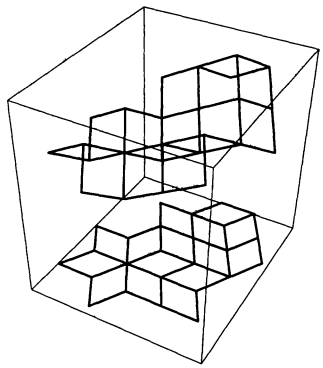


FIG. 2.2. Three-dimensional interpretation of lozenge tiling. If a region R can by tiled by lozenges, then the lozenge pattern lifts to the 2-skeleton of a cubical tiling of R^3, oriented diagonally to the plane of the lozenges.

There is a very direct geometric interpretation: think of the graph $\Gamma(L)$ as the 1-skeleton of a cubical tesselation of space, oriented so that cubes are on their corners: more precisely, so that the two endpoints of any path labeled *abc* are on the same vertical line. The 2-complex $\Gamma^2(L)$ is the union of the faces of the cubes. A lozenge in the plane is the orthogonal projection of a square face of a cube. Given a path π in the plane, arrange it (for notational purposes only) so the base point $*$ lies below the base point 1 of $\Gamma(L)$. Lift it edge by edge to a path in $\Gamma(L)$. When you make a complete circuit around π, you may or may not come back to the starting point in $\Gamma(L)$. The invariant $I(\pi) \in L$ is the ending vertex. This invariant of necessity lies in the kernel of the map $L \to A$, which is isomorphic to $\mathbf{Z}$: it can be described simply as the net rise in height.

If R can be tiled by lozenges, the tiling itself can be lifted, tile by tile, into $\Gamma^2(L)$, that is, into the 2-skeleton of the cubical tesselation. This gives another proof that the invariant $I(\pi)$ must be 1 if R can be tiled. In fact, if you look at a tiling by lozenges, you can imagine it so that it springs out at you in a three-dimensional picture.

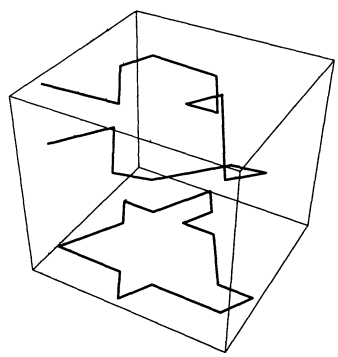


FIG. 2.3. Nontileable region. The region in the plane enclosed by the polygonal curve cannot be tiled by lozenges, since when it is lifted to the cubic network, it fails to close.

Algebraically, given the word representing π, the net rise in height is simply the sum of the exponents. The condition is that π heads at a bearing of 0°, 120°, or 240° the same length of time it heads at a bearing of 60°, 180°, or 300°.

This condition can be seen in an alternative way using a coloring argument. The triangles in the plane have an alternating coloring, with *abc* triangles colored white and *cba* triangles colored black. Each lozenge covers one triangle of each color—therefore, if R can be tiled, the number of white triangles must equal the number of black triangles. The difference in fact can be shown to be the net rise in height of α, as measured in main diagonals of cubes. The coloring consideration really gives a more elementary derivation that $I(\pi)$ must vanish for a tiling to be possible. However, this and related coloring arguments in general cannot give as much information as $I(\pi)$. One way to think of it is that coloring arguments are the abelian part of the group theory. If the group is abelian as in the present case, or more generally if the subgroup consisting of invariants $I(\pi)$ for *closed* paths is abelian, then that information is sufficient.

The algebraic condition that $I(\pi) = 1$, is not sufficient to guarantee a tiling by lozenges. There are curves π which go around nearly a full circle, with the lift in $\Gamma(L)$ rising considerably, and then instead of closing, they circle around another loop which brings them down to the starting height. If R could be tiled by lozenges, it could be divided into two regions by a fairly short path along edges of lozenges: but the rise in height for one side would be forced to be still positive, which would be a contradiction. We will return later to give a necessary and sufficient condition for a tiling by lozenges, along with a formula for a tiling if such exists.

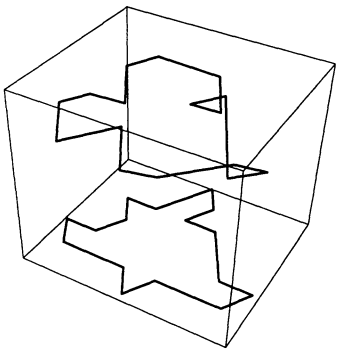


FIG. 2.4. Potentially tileable region. The boundary curve of this region lifts to a closed curve, so it meets the group-theoretic tiling condition. An actual tiling will be shown in 4.1, High lozenge tiling.

3. Tribone Tilings

Here is another example, for which other methods seem inadequate. I first heard this problem in an electronic mail inquiry from Carl W.Lee (ms.uky.edu!lee) in Kentucky.

> Last semester, a number of us here became interested in a combinatorial problem that was making the rounds. I'm sure you already have heard of it, and we heard a rumor that John Conway had solved it. It concerned a triangular array of dots. The problem was to pack in as many segments as possible, where each segment covered three adjacent dots in one of the three directions, and no two segments were allowed to touch. Is there any size configuration that admits a packing such that each dot is covered? Do you know anything about the status of this problem? Thanks in advance.

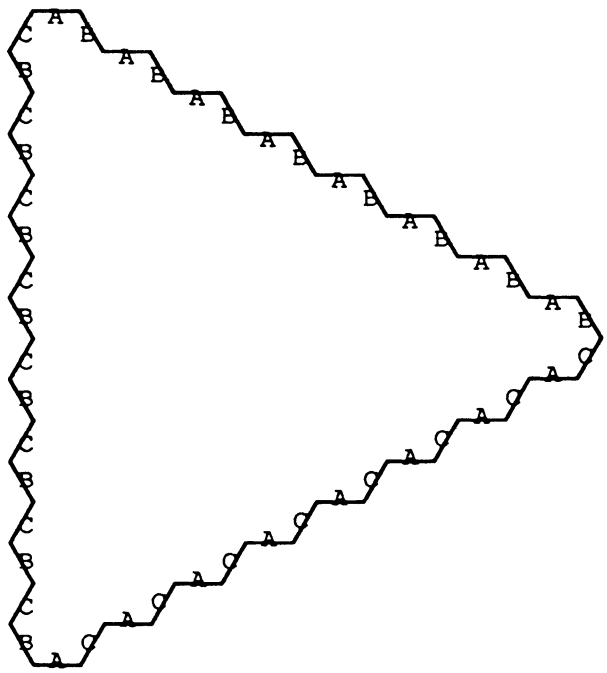


FIG. 3.1. Triangle of hexagons. A triangular array of hexagons, eight on a side. Can this be tiled by tribones?

I hadn't heard of it, but I asked Conway about it. We sat down together, and he worked it out.

This question can be alternately formulated in terms of a triangular array of hexagons. The problem is to show that one cannot tesselate the region using tiles made of three hexagons hooked linearly together. More generally, one can ask for the minimum number of holes left in an attempt to tile the region by these tiles.

If the region has side length n, then the number of hexagons is $n(n+1)/2$. A first, necessary condition is that n or $n+1$ is divisible by 3, that is, n is congruent to 0 or 2 mod 3. Note that if it is ever possible to solve the problem when n is congruent to 2 mod 3, one can extend the solution by adding a row of tiles along one side, to derive a solution for $n+1$.

Label each side in the hexagonal grid with an a, b, or c, according to the direction of the edge: a if it is parallel to the x axis, b if the angle from the x-axis to the edge (measured counterclockwise) is 60°, and c if this angle is 120°. Thus, the sides of every hexagon are labeled *abcabc*.

This labeling gives the 1-skeleton of the grid the structure of a group graph, where the group is

$$A = \langle a, b, c | a^2 = b^2 = c^2 = (abc)^2 = 1 \rangle.$$

The group is a group of isometries of the plane, generated by 180° revolutions about the centers of the edges; it also contains the 180° revolutions about the centers of the hexagons. The group A is sometimes called the $(2,2,2,2)$-group.

A path π in the 1-skeleton of the hexagonal grid now is determined by a word in the generators of A. We prefer to think of this in a slightly different way: π determines an element $\alpha(\pi)$ in the free product $F = \mathbf{Z}_2 * \mathbf{Z}_2 * \mathbf{Z}_2$. We are

particularly interested in closed paths, that is, elements of the kernel of $F \to A$. Unfortunately, this kernel is infinitely generated: it is a free group whose generators are given by arbitrary paths p_1, followed by a circuit around one of the three hexagons at the endpoint of p_1, followed by the p_1^{-1}.

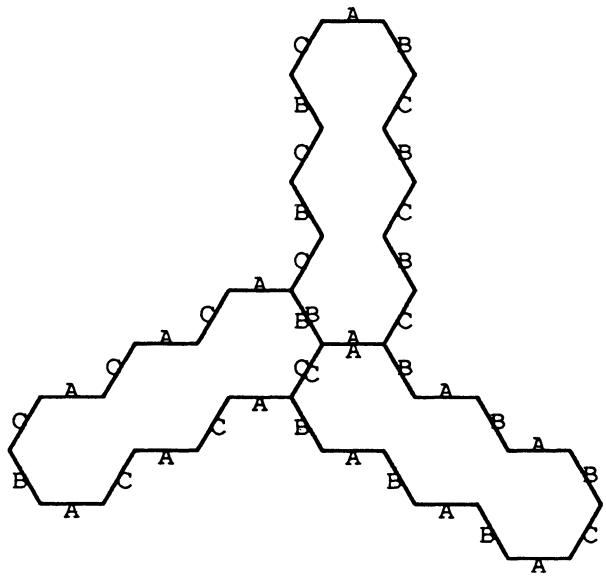


FIG. 3.2. Tribones in three orientations. There are three possible orientations for a tribone, in an array of hexagons. With our labeling convention, they are labeled in three different ways.

The standard tile, let us call it a *tribone*, can be laid in the plane in three different orientations. Circuits around the tribones in these three orientations trace out the elements

$$T_1 = (ab)^3 c (ab)^3 c$$

$$T_2 = (bc)^3 a (bc)^3 a$$

$$T_3 = (ca)^3 b (ca)^3 b.$$

If π is a simple closed circuit in the plane such that the region R bounded by π can be tiled by these tribones, then the image $I(\pi)$ of $\alpha(\pi)$ in the tribone group

$$T = \langle a, b, c | a^2 = b^2 = c^2 = T_1 = T_2 = T_3 = 1 \rangle$$

must be trivial.

The relation T_1 says that c conjugates $(ab)^3$ to its inverse. Observe that a and b also conjugate $(ab)^3$ to its inverse—in fact, this is already true in F. In other words, $(ab)^3$ generates a normal subgroup, and it commutes with every word of even length. Similarly, $(bc)^3$ and $(ca)^3$ generate normal subgroups. Together, the three elements generate a normal abelian subgroup J of T.

To form a picture of T, let us first look at the quotient group $T_0 = T/J = \langle a, b, c | a^2 = b^2 = c^2 = (ab)^3 = (bc)^3 = (ca)^3 = 1 \rangle$. The graph of T_0 can readily be constructed: take an infinite collection of three types of hexagons, with their edges labeled by the relations C_1, C_2 and C_3. These glue together to form a hexagonal pattern in the plane, where each vertex has one a edge, one b edge, and

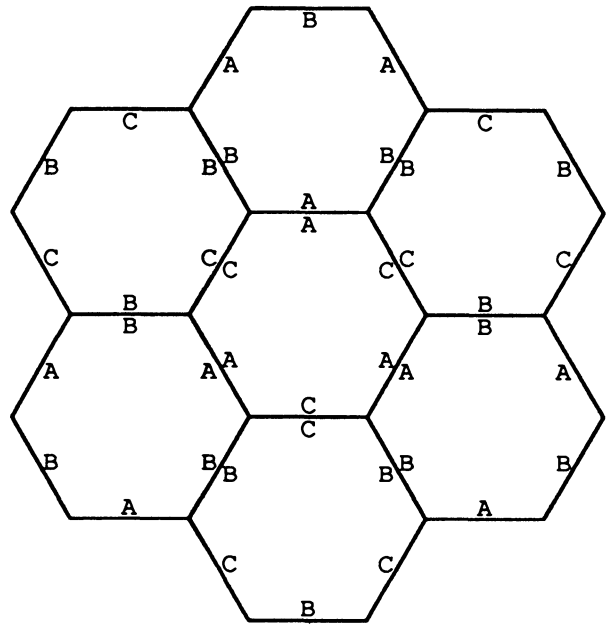


FIG. 3.3. Second hexagonal group. The group T_0 also has a graph isomorphic to the edges of a hexagonal tiling of the plane.

one c edge incident to it. The group T_0 acts faithfully as a group of isometries of the plane, generated by reflections in the edges of this hexagonal tiling: it is a triangle group. It is curious that even though the groups A and T_0 and the labeled graphs $\Gamma(A)$ and $\Gamma(T_0)$ are different, when the labels are stripped they become isomorphic.

If the region R can be tiled by tribones, then $\alpha(\pi)$ must map to the trivial element of T, so it maps to the trivial element of T_0. In our case, the region is a triangular array of hexagons, and its boundary can be taken as $\alpha(\pi) = (ab)^n(ca)^n(bc)^n$.

Obviously, if n is a multiple of 3, the image $I(\pi)/J$ in T_0 is trivial. In the other case, that n is 2 more than a multiple of 3, it is also trivial. This is easily seen by tracing out the curve in our array of hexagons, or by noticing that one can add additional tribones along one edge to form a triangular region with side length $n + 1$, which is a multiple of 3. Since we have pushed π only across tribones, $I(\pi)$ is the same for the two cases.

Since T_0 was not sufficient to detect the nontriviality of $I(\pi)$, we need to finish our job, and build a picture of T. First, look at the path in the graph of T_0 determined by the element T_1. Start at a vertex $*$ where the circuit $C_1 = ababab$ goes counterclockwise around a hexagon. Then T_1 goes counterclockwise around this hexagon, then along the c edge, clockwise around the C_1 hexagon through that vertex, and back along the c edge to close. In particular, the signed total of C_1-hexagons enclosed (counted according to degree of winding with counterclockwise circuits counted positively), is 0.

It is not hard to describe now the full group T, which is an extension of the form $J = \mathbf{Z}^3 \to T \to T_0$. We can interpret an element of T to be a vertex v in the graph of T_0, together with a path p from $*$ to v, subject to the equivalence relation that

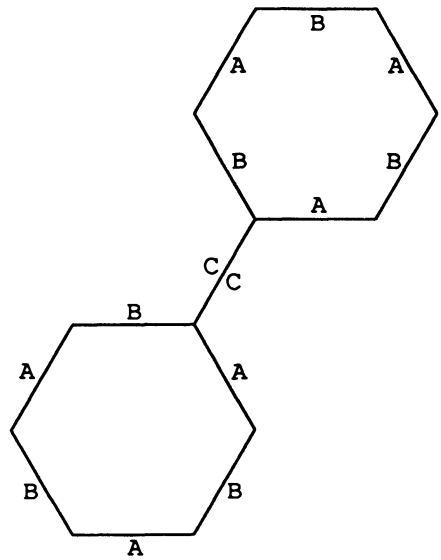


FIG. 3.4. Alternate image of tribone. By construction, the tribone relations are satisfied in the groups T and hence $T_0 = T/J$. This is the image of one of the tribone relators in the graph of the group T_0. Note how it encircles two AB-hexagons, once clockwise and once counterclockwise.

if q is another path from $*$ to v, then $p \sim q$ if the signed totals of C_1, C_2, and C_3 hexagons are all 0. (Of course, if we pick one path such as p from $*$ to v, then other paths from $*$ to v are determined by three arbitrary integers, which specify these signed totals.) With this definition, the relations T_i are obviously satisfied, hence the group so constructed is at least a quotient group of T. But we have already seen that the kernel J of the map $T \mapsto T_0$ is abelian, and generated by C_i. In the construction, this kernel is the free abelian group on the C_i, so it must in fact give T.

Once we know T, we can read $I(\pi)$ by inspection. As we saw, it suffices to consider the case $n = 3k$; the invariant is $C_1^k C_2^k C_3^k$, which is obviously not 1, so the tiling is impossible.

One can ask whether this method gives a lower bound on the number of holes one is forced to leave, in a partial tiling of R by tribones. To study this question, we should examine the subgroup K of T generated by elements of the form $I(\gamma)$, where γ is a path in the graph of A going from $*$ to some point v, circumnavigating a hexagon, and returning. In other words, K is the kernel of the map $T \to A$. Note that $\alpha(\gamma)$ has the form $gabcabcg^{-1}$, where g is arbitrary. In the group T_0, $abcabc$ acts as a translation. The conjugates of $abcabc$ in T_0 are translations in three different directions spaced at 120° angles, and the subgroup they generate is ismorphic to $\mathbf{Z}^2$. In K, there are actually an infinite number of different conjugates of $abcabc$: if g acts as a translation in T_0, then the commutator $gabcabcg^{-1}cbacba$ is trivial in T_0, but it might not be trivial in T: this path may enclose an arbitrary number m of hexagons of type C_1, and an equal number of type C_2 and C_3.

The subgroup K is therefore a nilpotent group, generated by $s = abcabc$, $t = bcabca$, and $u = C_1C_2C_3$, with presentation

$$K = \langle s, t, u \| [s, u] = [t, u] = 1, [s, t] = u^3 \rangle.$$

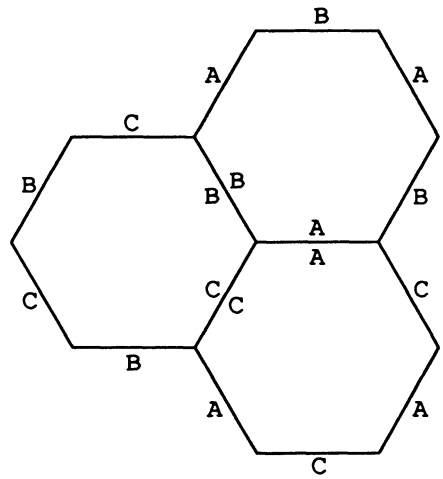


FIG. 3.5. Alternate image of a triangle. The triangle word $(ab)^n(ca)^n(bc)^n$ of size $n = 3m$ or $n = 3m + 2$ maps to the trivial element in T_0. In the diagram above, if $n = 3m$, trace the word starting at the center. If $n = 3m + 2$, start b from the center.

It is easy to check that every element of K is realized as $I(\pi)$, for some simple closed curve π in the plane.

Even though the invariants associated with triangular regions take larger and larger values in I, this does not give any information limiting the number of holes: for instance, three holes $g_i abcabcg_i^{-1}$ can yield u^k, for arbitrarily high k. In fact, it is possible to tesselate the triangular region of size n with tribones except for 1 hole, if $n \equiv 1(3)$, by placing the hole exactly in the middle, and then arranging concentric triangular layers of tribones around this hole. From these examples, tribone tilings with 3 holes are easily constructed when $n \equiv 0(3)$ or $2(3)$. It does give some information, however: in the case that $n \equiv 2(3)$ or $n \equiv 0(3)$, the conjugacy class changes ("increases") with n, which implies that the length of the minimum closed loop enclosing all the holes has to go to infinity with n. In the case $n \equiv 1(3)$, the conjugacy class of $I(\pi)$ is constant—since the region can always be tiled with a single hexagon missing, $I(\pi)$ is conjugate to $abcabc$. However, the actual word changes with n, which implies that the missing hole cannot be too close to the boundary. Perhaps a careful analysis would show that if there is a single hole, it must be exactly in the center of the triangle.

4. Dominoes and Lozenges Revisited

Conway's tiling groups are quite versatile, provided you can work out the group determined by the tiles. Even when (or perhaps especially when) the invariant $I(\pi)$ gives no information which could not have been easily obtained by other means, the geometric picture of the graph of the group can sometimes be exploited to give not just an algebraic criterion, but a precise geometric criterion for the existence of a tiling.

When G is a tiling group (with presentation given by a set of tiles), we define a measure of area in $\Gamma^2(G)$ to be the area defined by projection to the plane: the area of a 2-cell is the area of a corresponding tile. When the algebraic invariant $I(\pi)$ is 1, the curve π bounding R lifts to a closed $\tilde{\pi}$ in $\Gamma(G)$. We can ask, what is

the minimum area of a surface S in $\Gamma^2(G)$ with boundary $\bar{\pi}$? This area is necessarily at least as great as the area of R. If it is equal, then the images of the 2-cells of S must be disjoint, so that they form a tiling of R. There are several approaches which are sometimes successful for calculating this minimal area, but there is one particular situation when there is a really definitive solution: when $\Gamma^2(G)$ can be enlarged, by adding 3-cells, to make a contractable 3-manifold. In this situation, there is a "max flow min cut" principle which guarantees an efficient algorithm for finding a minimal surface.

Rather than going on with the general theory, we will illustrate this with two examples. First we revisit the lozenge question.

If R is a union of triangles in the plane, and if v and w are vertices in R, possibly on the boundary, define $d(v, w)$ to be the minimum length of a positively directed edge-path in R (possibly going on the boundary) joining v to w. This "distance" function d is not symmetric, since we cannot simply reverse an edge path. Any closed positively directed edge path has length a multiple of 3, so the $d(v, w)$ is defined modulo 3 independent of path. The three vertices of a triangle take the three distinct values modulo 3. If R is connected, it is always possible to find at least one positively directed path from v to w, so $d(v, w)$ is well-defined.

Consider the lifting of any tiling of R by lozenges to the cubical network, $\Gamma^2(L)$. This is determined by a height function $h(v)$ for the vertices v. We can choose the vertical scale so that h is integer-valued, and each edge of a lifted lozenge increases in height by 1; the edge of the triangular network covered by the lozenge lifts to a diagonal of a square, and decreases in height by 2. It follows that $h(w) - h(v) \geq d(v, w)$.

The boundary path π determines a unique height function h on its vertices, up to constants. This gives a necessary condition that R can be tiled: for any two vertices v and w on π, $h(w) - h(v) \geq d(v, w)$.

If π satisfies this necessary condition, then there is a unique maximally high lozenge tiling: define

$$h(x) = \min_{v \in \pi} \{d(v, w)\}.$$

To produce the actual tiling, place a lozenge so as to cover an edge where the height changes by 2. Since the three vertices of a triangle take distinct values moduluo 3, and since h increases by at most 1 along any edge, each triangle has exactly one edge where h changes by 2: therefore, the collection of lozenges is a tiling.

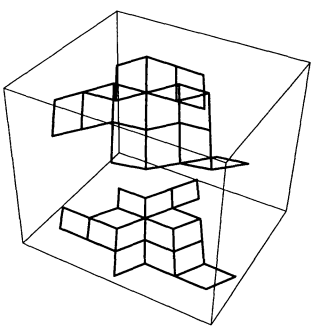


FIG. 4.1. High lozenge tiling. The "highest" lozenge tiling compatible with the boundary curve.

There is a simple algorithm for quickly computing h, and the tiling: rather than spell it out, we will describe the analogous algorithm for dominoes.

A closed path π in a square grid can be described by an element $\alpha(\pi)$ of the free group $F(x, y)$, which maps to the trivial element of the $A = \mathbf{Z}^2$. If the region R bounded by π can be filled with dominoes, then the image $I(\pi)$ of $\alpha(\pi)$ in the domino group $G = \langle x, y | xy^2 = y^2x, yx^2 = x^2y \rangle$ must be trivial.

What does the graph of G look like? We can construct a picture in $\mathbf{R}^3$, as follows. Fill the xy-plane with a black and white checkerboard pattern. Above the black square $[0, 1] \times [0, 1]$, construct a right-handed helix, joining $(0, 0, 0)$ by a line segment to $(0, 1, 1)$, to $(1, 1, 2)$, $(0, 1, 3)$, $(0, 0, 4)$, and so on: the x and y coordinates here marching forever around the boundary of the square, while the z coordinate increases by 1 each move. Similarly, $(0, 0, 0)$ is connected to $(0, 1, -1)$, etc. Construct a similar helix above each black square. Label each edge x or y, according to its image in the plane. Note that this creates left-handed helices above the white squares. The boundary of any domino in the plane lifts to a closed path in this graph we have constructed. Since the graph has a simply-transitive group of isometries, it is the graph of a group. Since it satisfies the domino relations, it is at least a quotient group of the domino group G. It is not hard (and strictly speaking, it is not logically necessary) to verify that this graph is indeed the graph of G.

The curve π lifts to a curve $\overline{\pi}$ in the graph of G. A convenient way to denote this, in the plane, is to record the height of the lift next to each vertex of π in the

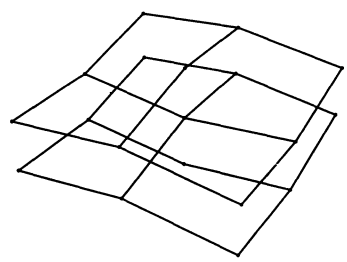


FIG. 4.2. The domino group. The graph of the domino group is a union of square helices over the squares of a checkerboard, alternating in handedness. A domino anywhere in the plane lifts to this graph, starting at any point. This illustration shows two coils of four neighboring helices.

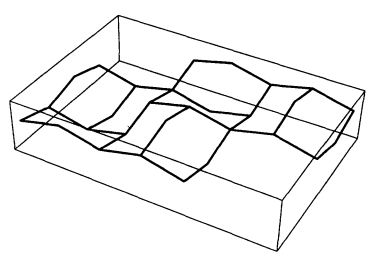


FIG. 4.3. Domino tiling. A tiling by 9 dominoes, lifted to the graph of the domino group.

plane. The rule is simple: one can start with 0 at some arbitrary vertex. Along any edge of π which has a black square to its left, the height increases by 1. Along any edge with a white square to its left, the height decreases by 1. A necessary condition that R can be filled with dominoes is that the height after traversing once around the curve is 0.

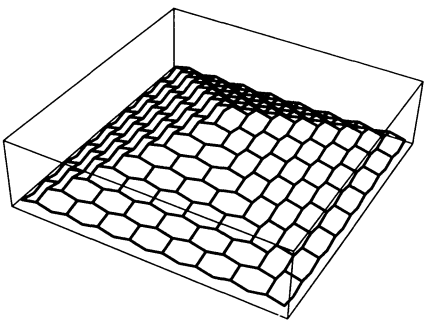


FIG. 4.4. Domino roof. This is the tiling which the algorithm yields, when applied to a 16×16 square grid. This is the tiling which has the highest lifting to the graph of the domino group of any tiling by dominoes.

There is a criterion and construction for a domino tiling, analogous to the construction for lozenges. Here is how the formula can be worked out, on a sheet of grid paper. Begin, as above, by labeling the height of each vertex of π. The heights consist of the integers in some interval, $[n, m]$. We will construct a height function on all vertices of R, beginning with $n + 1$, and working up. Suppose, inductively, that we have finished with all vertices of height less than or equal to k. For each vertex v of height k, and for each edge e leading from v which has a black square on its left, consider the second endpoint w of e. If the height of w has been previously defined, and if it is not greater than $k + 1$ leave it as is. If the height is defined and greater than $k + 1$, then a domino tiling is impossible: give up. Otherwise, define the height of w to be $k + 1$.

If this procedure reaches a successful conclusion, each edge of R has a difference of heights of its two endpoints of either 1 or 3. (Note that the height modulo 4 is determined by the point in the plane.) Erase all the edges whose endpoints have a difference of height of 3. What is left is a picture of a tiling by dominoes.

5. Triangles

Here is a related sequence of tiling problems which are resistant to direct attempts at general solution, but translate nicely into the realm of group theory.

Consider, again, a triangular array of dots, with N dots on each side. Is it possible to subdivide this array into disjoint triangular arrays of dots with M on each side? We suggest the reader indulge in experimentation with a few cases,

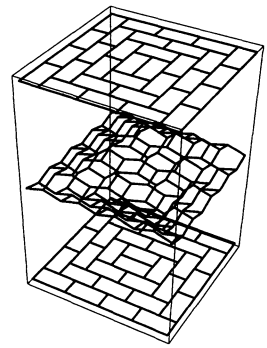


FIG. 4.5. Domino bubble. This illustration shows both the highest and the lowest tiling by dominoes of a standard checkerboard. They are isomorphic, differing only by a 90° rotation of the checkerboard (interchanging colors). The upper tiling is shown in the upper plane as well as the upper surface of the bubble, the lower tiling in the lower plane and the lower surface of the bubble. The bubble they form encloses the lift of any tiling by dominoes. Possible tilings are "like" Lipschitz functions in the square with Lipschitz constant 1, as measured in the Manhattan metric. The limits of domino tilings, lifted to the graph of the group, as the grid size goes to zero, are exactly such Lipschitz functions.

before reading further. For example, the cases $M = 2$ with N ranging from 2 to 12 are interesting.

...

As in the case of the tribones, this translates into a tiling problem: given a triangular array of hexagons with N hexagons per side, can one tile it by tiles T_M which are triangular arrays of hexagons M per side? We can express this with notation as in the case of tribones: label the edges of the underlying hexagonal tiling by a's, b's and c's. Given a path π in the plane, it is described by an element $\alpha(\pi)$ of $F = \langle a, b, c | a^2 = b^2 = c^2 = 1 \rangle$. If the region R bounded by π can be tiled by the copies of T_M, then the image $I(\pi)$ of $\alpha(\pi)$ is trivial in the group

$$G_M = \langle a, b, c | a^2 = b^2 = c^2 = 1, t_M = 1 \rangle,$$

were t_M represents the boundary curve of the tile T_M,

$$t_M = (ab)^M (ca)^M (bc)^M.$$

A parallelogram of hexagons with M hexagons on one side and $M + 1$ on the other can be tiled by two copies of T_M. This implies that $(ab)^M$ commutes with $(bc)^{M+1}$ and with $(ca)^{M+1}$, and so forth.

These relations imply that $(ab)^M$ commutes with $(bc)^{M(M+1)}$, and they also imply that $(ab)^{M+1}$ commutes with $(bc)^{M(M+1)}$. Combining these two facts, it follows that (ab) commutes with $(bc)^{M(M+1)}$. Geometrically, one can tile an $M \times M(M+1)$ parallelogram and an $M+1 \times M(M+1)$ parallelogram. Their difference is a $1 \times M(M+1)$ parallelogram: this can be tiled in a certain algebraic sense as the difference of the two.

It will simplify the picture at this point if we pass to the subgroups F^e and G_M^e generated by words of even length. Since all relations have even length, the wordlength modulo 2 describes a homomorphism of F and G_M to $\mathbf{Z}_2$, and these subgroups have index 2. The group F^e is the free group on 2 generators, but a more symmetric description is

$$F^e = \langle x, y, z | xyz = 1 \rangle,$$

where $x = ab$, $y = bc$, and $z = ca$. A presentation for the group G_M^e is obtained by adjoining relations coming from t_M to F^e: it requires two relations, one obtained by transcribing t_M directly, and the other transcribing the conjugate of t_M by an element of odd length. Using $t_M = 1$ and $bt_Mb = 1$, we obtain

$$G_M^e = \langle x, y, z | xyz = 1, x^M y^M z^M = 1, x^{-(M+1)} y^{-(M+1)} z^{-(M+1)} = 1 \rangle.$$

G_M^e has an interesting alternate generating set: $X = x^M$, $X' = x^{-(M+1)}$, together with Y, Y', Z and Z' defined similarly, clearly generate. We have already seen that X, Y, and Z commute with X', Y', and Z'.

The elements $s = X^{M+1}$, $t = Y^{M+1}$, and $u = Z^{M+1}$ commute with everything in G_M^e, so they generate a central subgroup J which is $\mathbf{Z}^3$ or a quotient. Let $G_M^0 = G_M^e/J$. We will analyze the structure of G_M^0, and from that construct G_M^e.

In G_M^0, X, Y, and Z satisfy relations

$$XYZ = 1, \qquad X^{M+1} = Y^{M+1} = Z^{M+1} = 1.$$

These relations describe the orientation-preserving $(M + 1, M + 1, M + 1)$ triangle group, which acts as a discrete group of isometries on the Euclidean plane if $M = 2$ and on the hyperbolic plane if $M > 2$. We have not checked that these generate *all* the relations on X, Y, and Z, but we immediately deduce that the subgroup H of G_M^0 generated by X, Y, and Z is a quotient of this triangle group. But there is a homomorphism f of the original group G_M to the full triangle group (including reflections), defined by sending a, b, and c to reflections in the sides of a $\pi/(M + 1)$, $\pi/(M + 1)$, $\pi/(M + 1)$ triangle. The relation $t_M = 1$ is satisfied, since in this group $(ab)^M = ba$ so that $(ab)^M(ca)^M(bc)^M = (ba)(ac)(cb) = 1$. Note that f sends X to ba, Y to ac and Z to cb, that is, to the standard generators of the $(M + 1, M + 1, M + 1)$ triangle group, and it sends s, t, and u to 0. Therefore, H is isomorphic to the orientable $(M + 1, M + 1, M + 1)$ triangle group.

A similar analysis shows that the subgroup H' generated by X', Y' and Z' is the orientable (M, M, M) triangle group. This group acts on the sphere, the Euclidean plane, or the hyperbolic plane when $M = 2$, $M = 3$, or $M \geq 4$. The analogous homomorphism f' maps G_M to the full (M, M, M) triangle group, mapping a, b, and c to the standard generators.

The two subgroups H and H' intersect (as seen from the effects of f and f') they generate G_M^0, and they commute with each other. Therefore, G_M^0 is the product $H \times H'$ of the two triangle groups.

Now we need to determine the kernel J of the quotient $G_M^e \to G_M^0$, and the structure of the central extension. As in the tribone case, we can do this geometrically, in terms of areas enclosed by curves. The graph Γ of the full $(M + 1, M + 1, M + 1)$ triangle group is formed from copies of three kinds of $2(M + 1)$-gons, with perimeters labeled $(ab)^M$, $(ca)^M$ and $(bc)^M$, with one of each kind meeting at each vertex. Arrange the orientation so that 1 is an "even" vertex that is, the a, b, and c edges emanating from 1 are in counterclockwise order. Then the relation t_M based at v encloses positively one copy of each type of polygon, while the conjugate bt_Mb encloses negatively one copy of each type of polygon.

Similarly, the graph Γ' of the full (M, M, M) triangle group is made from three kinds of $2M$-gons. Starting at the 1, which we suppose is an even vertex, the relation t_M encloses positively one copy of each type of polygon, while bt_Mb encloses negatively one copy of each. However, in the case $M = 2$, the entire graph is finite: it is the 1-skeleton of a cube, and the number of polygons enclosed by a curve is well-defined only modulo 2.

First let's deal with the case $M > 2$. We can define an extension K of G_M^0 as an equivalence relation on elements of F^e, as follows. An element g of F^e determines paths $p(g)$ in Γ and $p'(g)$ in Γ'. We define g to be equivalent to h if $p(g)$ ends at the same point as $p(h)$, $p'(g)$ ends at the same point as $p'(h)$, and if the closed loop $p(g)p^{-1}(h)$ encloses the same numbers of ab-polygons, bc-polygons, and ca-polygons as $p'(g)p'^{-1}(h)$.

In particular, an element of the kernel of the map of K to $H \times H'$ maps to closed loops in both pictures, and is determined by the triple of differences of the number of polygons enclosed. The elements s, t and u map to $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. It follows that $K = G_M^e$, and $J = \mathbf{Z}^3$ (provided $M > 2$).

The boundary of the size N triangle T_N can be described by the element $t_N = (ab)^N(ca)^N(bc)^N$. The path $p(t_N)$ in Γ closes only when N is 0 or -1 mod $M + 1$, while the path $p'(t_N)$ closes only when N is 0 or -1 mod M. Since M and $M + 1$ are relatively prime, there are four solutions modulo $M(M + 1)$: $0, M$, $M^2 - 1, -1$. For values of N satisfying one of these congruence conditions, the invariant in G_M^0 is 0, so the invariant is in J; it is a positive multiple of $(1, 1, 1)$ in all but the trivial case $N = M$.

THEOREM (Conway). *When $N > M > 2$, the triangular array T_N of hexagons cannot be tiled by T_M's.*

This analysis has an interesting variation case $M = 2$. Given two elements g and h of F^e, we can define them to be equivalent if $p(g)$ and $p(h)$ have the same endpoints, $p'(g)$ and $p'(h)$ have the same endpoints, and if the numbers of polygons of the three types enclosed by the path $p(g)p(h)^{-1}$ is a multiple k of $(1, 1, 1)$ which has the same parity as the number of polygons enclosed by $p'(g)p'(h)^{-1}$. This defines a central extension of $H \times H'$ by $\mathbf{Z}^3$ modulo the subgroup generated by $s^2t^2u^2 = 1$. To justify that this group is in fact G_2^e, we must prove that $s^2t^2u^2 = (ab)^{12}(ca)^{12}(bc)^{12} = 1$ in this group, or even better, that it is possible to tile T_{12}. Such a tiling can be found fairly easily—see FIGURE 5.1, the 12-stack by 2-stacks.

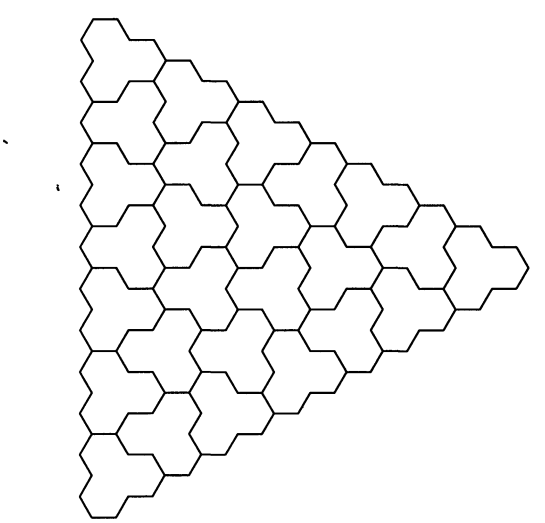


FIG. 5.1. The 12-stack by 2-stacks. The triangle T_{12} can be tiled by T_2's.

The computation of the mod 2 invariant for tilings by T_2's can be rather annoying when done directly. However, there is a neat trick, which enables one to see this invariant geometrically: most regions which have a multiple of 3 hexagons can be tiled easily by T_2's along with tribones. The boundary *abababcababababc* of a tribone maps to closed paths in both Γ and Γ'. In Γ, it encloses a net of 0 of each type of hexagon, as we saw before. In Γ', this curve winds counterclockwise 1.5 revolutions about an *ab*-face of the cube, goes down a *c*-edge to the opposite face, winds 1.5 revolutions counterclockwise (with respect to the orientation of the square), and goes up again to close. It is therefore equivalent, in terms of which kinds of squares it encloses, to *abcabc*, which is an odd multiple of (1, 1, 1).

Therefore, if a region can be tiled with a collection of T_2's together with an odd number of tribones, it cannot be tiled with T_2's. For $0 < N < 12$, only for the values 2, 3, 5, 6, 8, 9, 11 is the number of tiles a multiple of 3. One quickly finds that in the cases T_3, T_5, T_6 and T_8 there is a tiling by one tribone and the rest T_2's, while T_2, T_9, and T_{11} can be tiled.

Given any tiling or partial tiling of T_k, with $k > 1$, it can be extended to a tiling or partial tiling of T_{k+12} by adding a $12 \times k$ parallelogram, together with a T_{12}. The $12 \times k$ parallelogram can be tiled by subdividing into 2×6 and 3×6 parallelograms.

THEOREM (Conway). *A triangular array T_k of hexagons can be tiled by T_2's if and only if k is congruent to* 0, 2, 9, *or* 11 *modulo* 12.

REFERENCE

[CoLa] J. H. Conway and J. C. Lagarias, *Tiling with Polyominoes and Combinatorial Group Theory*.

Mathematical Writing



USING WRITING TO TEACH MATHEMATICS

Andrew Sterrett, Editor

...the day will come, I believe, when the value of writing to learn will be universally acknowledged

Rueben Hersh in
Writing to Learn Mathematics and Science

Need help in getting started as an individual or as a member of a department facing a Writing Across the Curriculum requirement? Learn how others have made use of student assistants, both undergraduate and graduate, in ways that benefit students and faculty members alike. Read that feedback from student journals provide early-warning signals for instructors, as well as help students clarify their own thought processes.

This collection of essays is an outgrowth of the widespread interest shown in sessions of contributed papers on writing given at the 1988 and 1989 Annual Meetings of the MAA. Many of the 30 authors of the essays included in this volume participated in those sessions and each has considerable experience in requiring students to write about mathematics.

Included in this volume are essays that:

- remind professors how frequently mathematicians, regardless of their careers, are asked to write, ("Mathematicians Write; Mathematics Students Should Too"),
- provide a theoretical framework by which to assess writing assignments ("Writing for Educational Objectives in a Calculus Course"), and
- give practical examples of assignments that work ("Writing in Mathematics: A Plethora of Possibilities").

This source book is filled with practical suggestions. It will enhance the comprehension that your students have of mathematics.

160 pp., Paperback, 1990,
ISBN-0-88385-066-4

List $12.50

Catalog Number NTE-16

ORDER FROM

The Mathematical Association of America

1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Springer for Geometry

R.S. Millman, Wright State University, Dayton, OH; and **G.D. Parker**, Southern Illinois University, Carbondale, IL

Geometry

A Metric Approach with Models
Second Edition

Imparts a real feeling for Euclidean and non-Euclidean (in particular, hyperbolic) geometry. Intended as a rigorous first course, the book introduces and develops the various axioms slowly, and then, in a departure from other texts, continually illustrates the major definitions and axioms with two or three models, enabling the reader to picture the idea more clearly. Over 700 problems and 250 figures are included in this revised second edition which has been expanded to include a selction of expositiory exercises.

Topics covered include: the fundamentals of neutral (absolute) geometry; the theory of parallels; hyperbolic geometry; classical Euclidean geometry; proof of the existence of an area function; the cut and reassemble theory of Bolyai; and the classification of isometries of a neutral geometry. Additionally, the authors have designed software with computational problems to accompany the text. This software may be obtained from George Parker.

1990/app. 345 pp., 261 illus./Hardcover
$49.00 (tent.)/ISBN 0-387-97412-1
Undergraduate Texts in Mathematics

S. Lang, Yale University, New Haven, CT; and **G. Murrow,** Phillips Academy, Andover, MA and Harvard School, Los Angeles, CA

Geometry

A High School Course
Second Edition

"There is a freshness about the exercise sections that could help inspire many a jaded teacher...on the one hand, rigorous and coherent, but on the other it is not 'pretentiously systematized'... Here we find synthetic geometry (á la Euclid), vectors, co-ordinates, and transformations all in their proper place. The mix is stimulating and avoids burdening the student with the philosophical idealism of some text-book writers." **— The Mathematical Gazette**

1988/394 pp., 577 illus./Hardcover/$39.00/
ISBN 0-387-96654-4

M. Berger, Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France

Geometry I and Geometry II

Translated from the French by
M. Cole and S. Levy

From the reviews:
"Berger's two volumes constitute a detailed introduction to classical geometry from the modern-group theoretic point of view...Berger offers a perspective on elementary plane geometry and looks ahead to advanced topics such as algebraic topology, Lie groups, and algebraic geometry. Illustrations and application of the material in art, architecture, engineering, crystallography, and cartography keep the presentation lively, and Berger makes various asides to amplify the connections between geometry and other parts of mathematics."
— CHOICE

Geometry I
1987/427 pp., 426 illus./Softcover/$39.80
ISBN 0-387-11658-3
Universitext

Geometry II
1987/406 pp., 364 illus./Softcover/$39.80
ISBN 0-387-17015-4
Universitext

Three Easy Ways to Order:

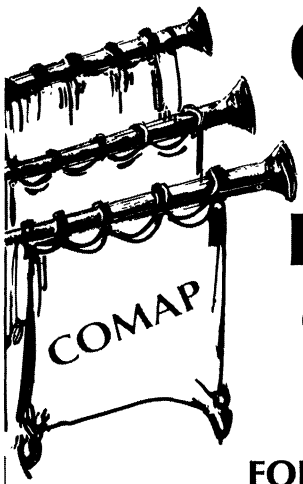
- **Call**: **Toll-Free 1-800-SPRINGER**
 in NJ call 201-348-4033 (8:30 - 4:30 PM EST).
 Your reference number is **S600.**
- **Write**: Send payment plus $2.50 for postage and handling to:
 Springer-Verlag New York, Inc.
 Attn: S. Klamkin - Dept. S600
 175 Fifth Avenue
 New York, NY 10010.
- **Visit**: Your local technical bookstore.

Order Today!
Call or write for examination copies!

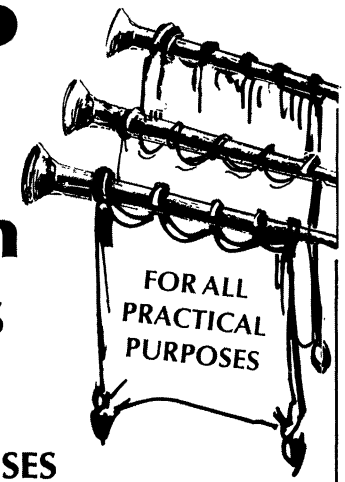


Springer-Verlag

New York • Berlin • Heidelberg • Vienna • London • Paris • Tokyo • Hong Kong • Barcelona



COMAP
The Revolution Continues



Presenting the Second Edition . . .

FOR ALL PRACTICAL PURPOSES
Introduction to Contemporary Mathematics
Second Edition

The Consortium for Mathematics and Its Applications (COMAP), Inc.

The book that inspired a revolution in liberal arts mathematics is back—in a new, even more effective Second Edition.

The first edition transformed the mathematics classroom by communicating the real-life significance—even excitement—of mathematics to students. It set aside rote computation and placed in its stead a vision of the power and elegance of math—*as applied to today's society, its issues, and its problems.* And these applications were fully integrated into the text—not added on as asides.

All this is continued in the Second Edition.

But how is this new version even stronger? There are more examples ★ twice as many exercises—with accompanying answers ★ more boxes focused on the human side of math ★ new optional sections for maximum teaching flexibility ★ and a great deal more.

Available November 1990.
Hardbound, 2115-5,
approx. 620 pages, 548 illustrations
Instructor's Manual (2180-5)
available free upon adoption.

Accompanying Annenberg/CPB *video-cassettes*—call 1-800-LEARNER for more information.

Transparency Masters and a *Telecourse Guide* are also available. Write to us at the address below for more information.

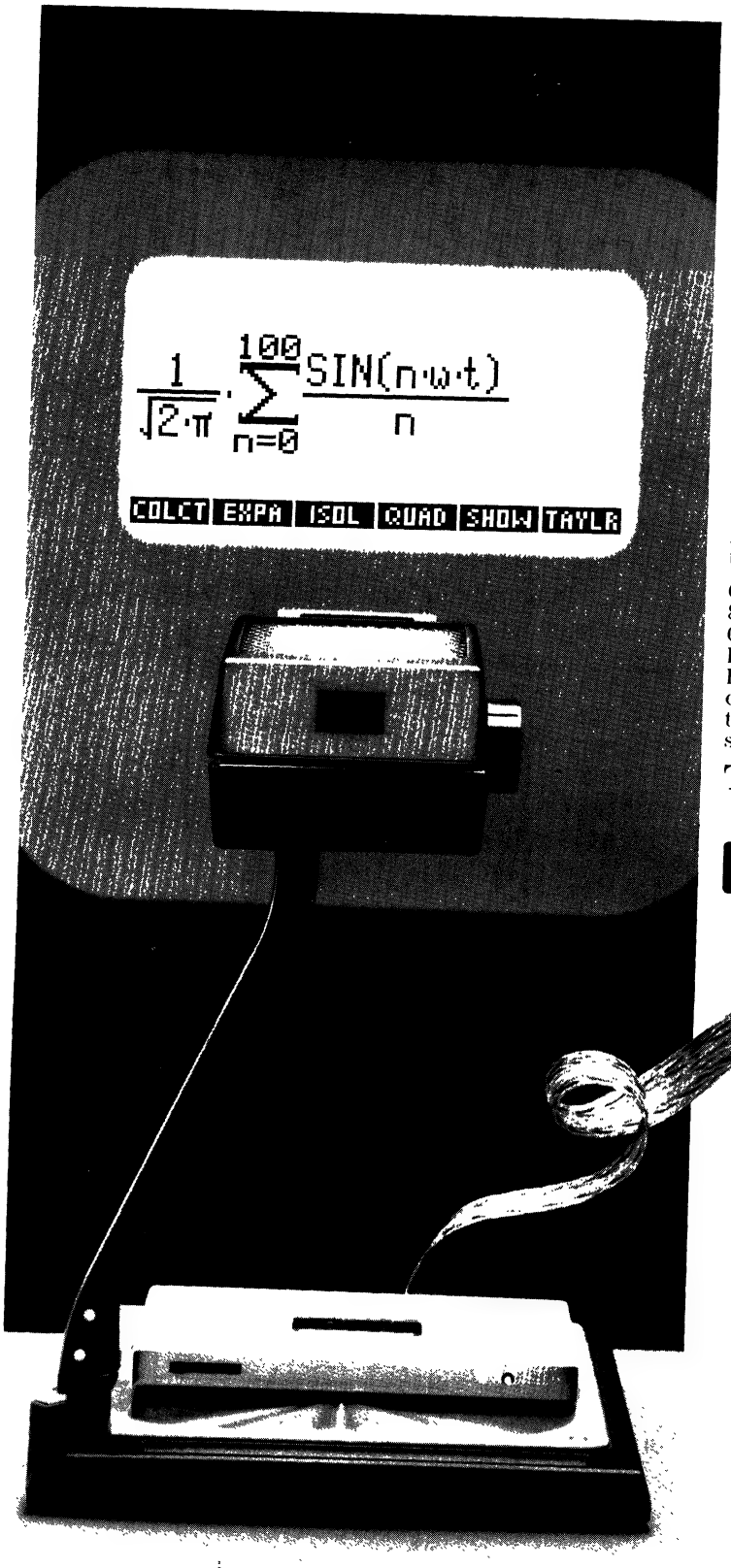
Request your examination copy of the new Second Edition of **For All Practical Purposes** today! Just write to us on your school letterhead, stating course name and number, annual enrollment, adoption decision date, and present text in use. Send it to:

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010



W. H. FREEMAN AND COMPANY
The book publishing arm of Scientific American

Help your students discover more meaningful relationships.



Free classroom display device with purchase of 30 calculators.

Showing is much more powerful than telling. So we've developed special classroom displays for our two most advanced calculators.

The new scientific expandable HP 48SX calculator and the cost-effective HP 28S are designed to put your students on the cutting edge of calculus and engineering. With more built-in functions and graphics solutions than any other calculator.

If your department or students purchase 30 HP 48SX or HP 28S calculators (or a mix of both), we'll give you a free HP 48SX and plug-in classroom display (an $850 retail value). Or a classroom display version of the HP 28S (a $500 retail value). And free teaching materials in the bargain.

Call us at **(503) 757-2004** between 8am and 3pm PDT for more details. Or write: Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330. Offer ends October 31, 1990, and applies to instructors at college and high school levels only.

There is a better way.

HEWLETT PACKARD

© 1990 Hewlett-Packard Company PG12005

When was the last time a computer program helped you think about mathematics?

Call or write for our latest software catalog.
For PC or Macintosh.

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

Enhance Your Math Courses with Calculator Manuals from HBJ

CALCULATOR ENHANCEMENT FOR
SINGLE-VARIABLE CALCULUS
Preliminary Edition
JAMES NICHOLSON
Consultant: JOHN KENNELY
Both of Clemson University
Paperbound, 1991. **Just Published!**

CALCULATOR ENHANCEMENT FOR
MULTIVARIABLE CALCULUS
Preliminary Edition
JAMES RENEKE, *Clemson University*
Paperbound, 1991. **Just Published!**

CALCULATOR ENHANCEMENT FOR
LINEAR ALGEBRA
Preliminary Edition
DON LA TORRE, *Clemson University*
Paperbound, 1991. **Just Published!**

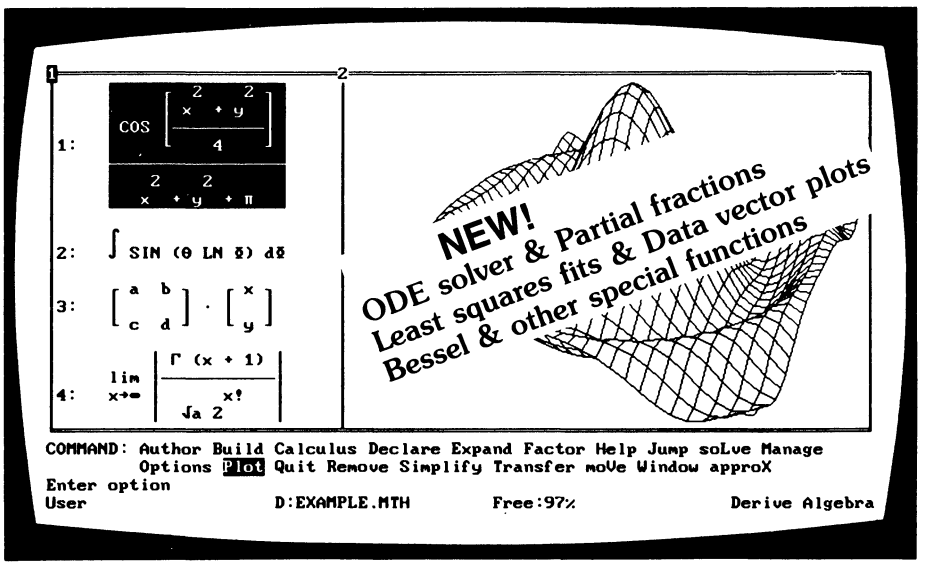
CALCULATOR ENHANCEMENT FOR
DIFFERENTIAL EQUATIONS
Preliminary Edition
THOMAS G. PROCTOR, *Clemson University*
Paperbound, 1991. **Just Published!**

CALCULATOR ENHANCEMENT FOR
INTRODUCTORY STATISTICS
Preliminary Edition
IRIS B. FETTA, *Clemson University*
Paperbound, 1991. **Just Published!**

HBJ **Harcourt Brace Jovanovich, Inc.**
College Sales Office
7555 Caldwell Avenue
Chicago, IL 60648
(708) 647-8822

DERIVE

A Mathematical Assistant



2000 years of mathematical knowledge on a disk

DERIVE, **the successor to muMATH, is a powerful computer algebra system for your PC compatible computer that provides the following capabilities:**

- **Exact and approximate arithmetic to thousands of digits**
- **Equations, complex numbers, trigonometry, calculus, vectors, and matrices**
- **2D and 3D function plotting with zooming capability**
- **MDA, CGA, EGA, VGA, and Hercules graphics and text support**
- **Attractive 2D mathematical display of formulas**
- **Easy to use menu-driven interface with on-line help**
- **Ideal for engineers, scientists, students and teachers**
- **$200 plus shipping: Call or write for information.**

System requirements: IBM PC or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (360K) or a 3½ inch (760K) diskette drive. **Or** NEC PC-9801 or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (640K) diskette drive.

DERIVE and muMATH are trademarks of Soft Warehouse, Inc. Hercules is a trademark of Hercules Computer Technology, Inc. IBM is a registered trademark of International Business Machines Corp. MS-DOS is a registered trademark of Microsoft Corp. NEC is a registered trademark of Nippon Electric Company.

Soft Warehouse INC
3615 Harding Avenue, Suite 505 • Honolulu, HI 96816
(808) 734-5801 after noon PST
Handcrafted software for the mind.

©1988 Soft Warehouse, Inc.

Addison-Wesley gives your students the best of math

Differential Equations: A Modeling Approach

Frank R. Giordano and Maurice D. Weir, *U.S. Military Academy and U.S. Naval Postgraduate School, respectively*

Now your students can clearly see how differential equations apply to realistic problems. This new text by Giordano and Weir focuses on mathematical modeling, showing students a natural progression of related ideas, and offering motivation to take the next step along the way. Unlike other texts that use a "solutions technique" approach, **Differential Equations: A Modeling Approach** puts into perspective how the various solutions techniques students learn actually fit together — the "why" as well as the "how-to."

Discrete Algorithmic Mathematics

Stephen B. Maurer and Anthony Ralston, *Swarthmore College and SUNY-Buffalo, respectively*

Turn to Maurer and Ralston for discrete mathematics with a basic difference: a unique algorithmic approach. The authors present discrete mathematics as a unified topic, by viewing it through an algorithmic "prism." A wide range of applications, flexible coverage of core discrete math topics, and extensive coverage of linear algebra also add to this text's substantial value.

Addison-Wesley
1 Jacob Way • Reading, MA • 01867
(617) 944-3700

From the publisher that's serious about math

ALSO NEW FOR '91!

Developmental Mathematics

Basic Mathematics, Sixth Edition
Introductory Algebra, Sixth Edition
Intermediate Algebra, Sixth Edition
Mervin L. Keedy and Marvin L. Bittinger, *Purdue University and Indiana University-Purdue University at Indianapolis, respectively*

Concepts and Applications of Intermediate Algebra, Second Edition
Marvin L. Bittinger, Mervin L. Keedy, and David Ellenbogen, *Indiana University-Purdue University at Indianapolis, Purdue University, and St. Michael's College, respectively*

Elementary Algebra
Intermediate Algebra
Mark Dugopolski, *Southeastern Louisiana University*

Statistics

Introductory Statistics, Third Edition
Neil A. Weiss and Matthew Hassett, *both of Arizona State University*

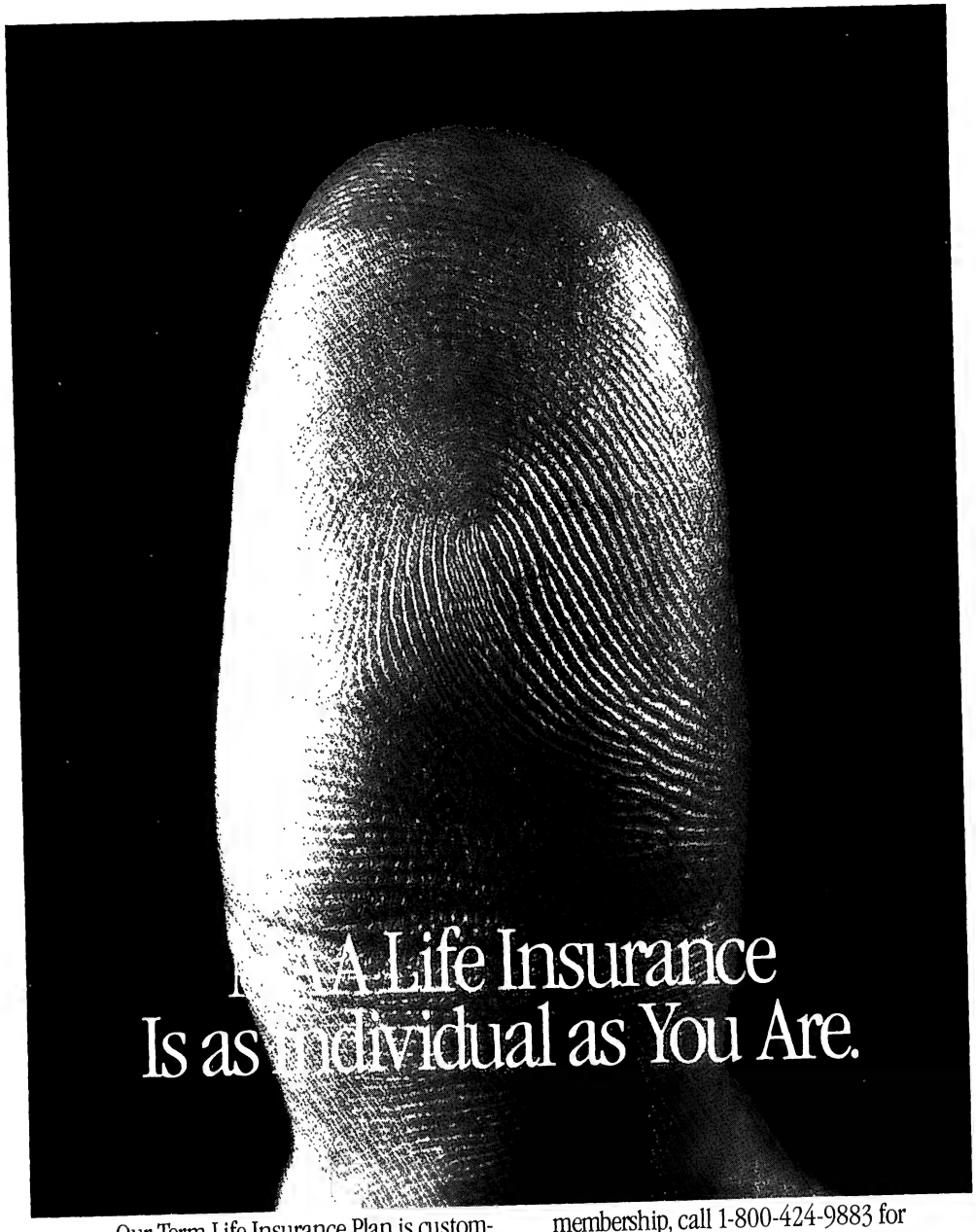
Statistics Methods for Business and Economics, Fourth Edition
Donald L. Harnett and Ashok Soni, *both of Indiana University*

OTHER TITLES OF INTEREST

Precalculus: Functions and Graphs
A. Robert Marshall, *Cuesta College*

Mathematics for Elementary School Teachers: A Problem Solving Approach
Joseph Newmark, *CUNY College of Staten Island*

Calculus: A Complete Course
Robert A. Adams, *University of British Columbia*



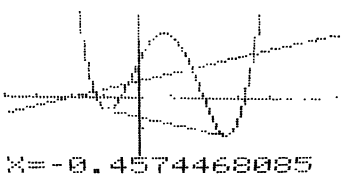
Our Term Life Insurance Plan is custom-designed for members of our profession. What's more, each policy can then be tailored to suit your individual needs. As these needs change, so can the policy—and it can stay with you no matter how often you change jobs.

Our group purchasing power helped us to negotiate top quality insurance, at a very low price. To take advantage of this benefit of membership, call 1-800-424-9883 for further details (in Washington, D.C. call 457-6820).

MAA INSURANCE
Designed by Members.
For Members.

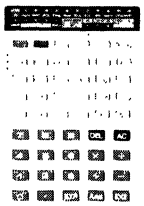
The MAA Life Plan is underwritten by Connecticut General Life Insurance Company, A CIGNA Company, Hartford, Connecticut 06152 on form number GM 3000.

The Power to Project Your Ideas



Introducing the OH-7000G Projectable Graphic Calculator.

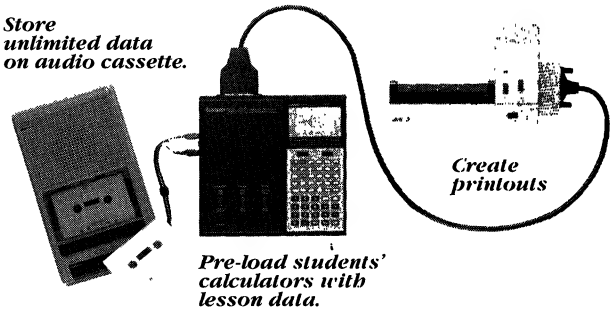
Casio has taken one of its most powerful graphic calculators and turned it into the most powerful **teaching tool** of its kind. The OH-7000G features a transparent, projectable screen so that teachers can lead their class just by using a standard overhead projector.



OH-7000G

Since inventing the graphic calculator in 1985, Casio has created an entire family of exciting classroom products. Five different graphic calculator models are available, each offering **hundreds of powerful, built-in functions**.

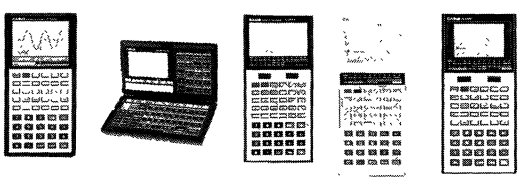
With our **printers**, you can create "hard copy" of graphs and calculations for photocopying onto test papers or transparencies. The **FA-80 Data Interface** lets you store unlimited data on audio cassettes.



The OH-7000G lets the instructor lead the class via a standard overhead projector.

The possibilities are endless. That's why the **NCTM curriculum** calls for the use of graphic calculators. Many **new textbooks** are being published especially for use with this technology.

So, take a bold, new step and enter Casio's "Electronic Classroom" – a place where the solution is always clear.



CASIO

Creators of the Graphic Calculator

570 Mt. Pleasant Avenue, Dover, New Jersey 0780

For more information about Casio's family of graphic calculators, and to place an order, call one of the educational distributors listed below.

ABM
800-234-3396
(In TX 214-869-0721)
(In MO 816-861-1113)
(In CA 213-946-6680)

Allied Communications
800-843-3660
(In PA 800-843-3660)

Allied National
800-999-8099
(In MI 313-543-1232)

Calculators Inc.
800-533-9921
(In MN 800-533-9921)

Carolina Wholesale
800-521-4600
(In NC 704-598-8101)

Colborn School Supply
800-548-7031
(In CO 303-778-1220)
(In MT 406-245-3158)

Copco
800-446-7021
(In OH 800-589-3006)

Dale Seymour Publications
800-872-1100
(In CA 800-222-0766)

Douglas Stewart
800-356-8370
(In WI 608-221-1155)

E.A.I.
800-272-0272
(In NJ 201-891-9466)

El Dorado Trading
800-227-8292
(In CA 415-494-1995)

Entertainment Marketing
800-347-2001
(In TX 713-995-4433)

Global Products
800-633-0633
(In IL 708-397-4944)

Longino Distributors
800-633-6224
(In NC 704-873-3282)

NASCO
800-558-9595
(In WI 414-563-2446)

Nemesis Distributing
800-940-7407
(In FL 305-477-8822)

Servco Pacific
(In HI 808-841-7566)

Tam's
800-421-5188
(In CA 800-244-5624)

Taylor Electric
800-558-6970
(In WI 800-242-8940)

Techline
800-777-3635
(In VA 703-389-0857)

Valley Business Machines
(In UT 801-969-6303)

Vistatech
800-847-9851
(In NY 212-254-9851)
(In CA 213-602-0277)

CASIO®

PLANT THE SEEDS TO HIGHER LEARNING . . .

Being first is almost second nature at McGraw-Hill. Whether your students are enrolled in developmental math, precalculus, or calculus, McGraw-Hill's texts will inspire your students to learn.

NEW FOR 1991

The Streeter Series

BASIC MATHEMATICAL SKILLS WITH GEOMETRY
Streeter/Hutchison/Hoelzle

BEGINNING ALGEBRA, Form B
Streeter/Hutchison/Hoelzle

INTERMEDIATE ALGEBRA, Form B
Streeter/Hutchison/Hoelzle

COLLEGE ALGEBRA WITH TRIGONOMETRY
Rees/Sparks/Rees

PRECALCULUS: A Bridge to Calculus
Runyan/Runyan

EXPLORATIONS IN CALCULUS WITH A COMPUTER ALGEBRA SYSTEM
Small/Hosack

DISCOVERING CALCULUS WITH THE HP-28s AND THE HP-48sx
Smith/Minton

DISCRETE MATH AND ITS APPLICATIONS, 2/e
Rosen

DIFFERENTIAL EQUATIONS WITH APPLICATIONS, 2/e
Simmons

INTRODUCTION TO PARTIAL DIFFERENTIAL EQUATIONS, 2/e
Pinsky

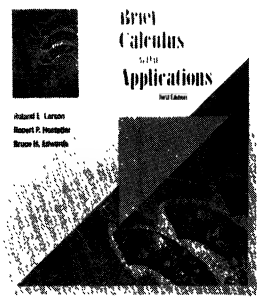
FUNCTIONAL ANALYSIS, 2/e
Rudin

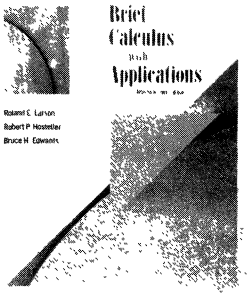
MATRIX THEORY
Goldberg

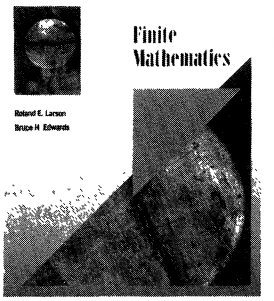
BRIDGE TO ABSTRACT MATHEMATICS, 2/e
Morash

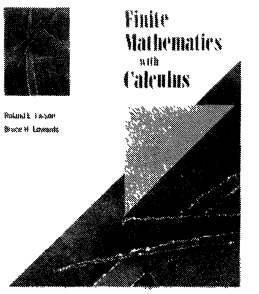
For more information, please contact your local McGraw-Hill representative or write: McGraw-Hill College Division, Comp Processing and Control, P.O. Box 448, Hightstown, NJ 08520.

Launching Another Year of Excellence: D. C. Heath Mathematics in 1991









*The popular **Brief Calculus**—now part of a remarkable series!*

Brief Calculus with Applications
Third Edition
1991 Cloth 812 pages

Brief Calculus with Applications
Alternate Third Edition
1991 Cloth 648 pages

Roland E. Larson and **Robert P. Hostetler**, both of The Pennsylvania State University, The Behrend College
Bruce H. Edwards, University of Florida

Finite Mathematics
1991 Cloth 500 pages

Finite Mathematics with Calculus
1991 Cloth 918 pages

Roland E. Larson, The Pennsylvania State University, The Behrend College
Bruce H. Edwards, University of Florida

Also new for 1991!

Understandable Statistics
Fourth Edition
Charles Henry Brase, Regis College
Corrinne Pellillo Brase, Arapahoe Community College
1991 Cloth 704 pages est.

Elementary Linear Algebra
Second Edition
Roland E. Larson, The Pennsylvania State University, The Behrend College
Bruce H. Edwards, University of Florida
1991 Cloth 680 pages

Fun and Games
A Text on Game Theory
Ken Binmore, University of Michigan
1991 Cloth 440 pages est.

Mathematics for Elementary Teachers: A Balanced Approach
Second Edition
Eugene F. Krause, University of Michigan
1991 Cloth 928 pages est.

Arithmetic for College Students
Sixth Edition
D. Franklin Wright, Cerritos College
1991 Cloth 656 pages est.

Basic Mathematics
Third Edition
David Novak, Simmons College
1991 Paper 592 pages est.

Recently published

Calculus with Analytic Geometry
Fourth Edition
Larson/Hostetler/Edwards
1990 Cloth 1083 pages

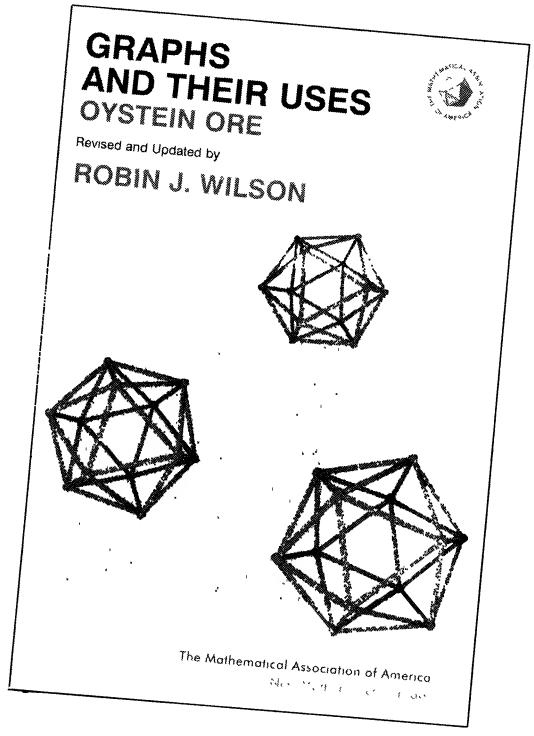
Calculus with Analytic Geometry
Alternate Fourth Edition
Larson/Hostetler
1990 Cloth 1113 pages

For more information, call us toll-free: **1-800-235-3565**.

D. C. Heath and Company

College Division • 125 Spring Street • Lexington, MA 02173

Revised and Updated



GRAPHS AND THEIR USES

Oystein Ore
Revised and updated by Robin J. Wilson

In 1963 Oystein Ore wrote this classic volume, which was published in the New Mathematical Library Series. This elegant book has provided students and teachers with an excellent introduction to the field of graph theory for close to thirty years. Robin Wilson's revision adds strength to the book by updating the terminology and notation bringing them in line with contemporary usage. Wilson has added new material on interval graphs, the travelling salesman problem, bracing frameworks, shortest route problems, and coloring maps on surfaces. Most of the diagrams in the book have been redrawn.

The book shows us how the theory of graphs may be employed for diverse purposes such as recording a baseball league's games at every stage of the season; representing a biological breeding experiment; scheduling an efficient chess tournament; determining the cheapest way of building a railroad net; solving some ancient puzzles; analyzing winning (or losing) positions in certain games. The reader studying these applications will become increasingly interested in the properties and applications of graphs to mathematical problems. After presenting and proving Euler's formula for polyhedra, this book culminates in a discussion of the famous Four-Color Conjecture.

Robin Wilson says in the Introduction to the revised edition, "I have always regarded Ore's text as a classic, and working on this second edition has served to reinforce this view. It is my hope that this edition will enable a new generation of readers to derive as much pleasure from Ore's book as my generation did in the 1960s and 1970s."

160 pp., Paper, 1990, ISBN 0-88385-635-2

List $14.50 MAA Member: $12.50

Catalog Number NML-34

ORDER FROM

The Mathematical Association of America

1529 Eighteenth Street, N.W.
Washington, D.C. 20036

Contents

(ISSN 0002-9890)

Introduction 649

Convexity
by Marcel Berger 650

What is Geometry?
by Shiing-Shen Chern 679

A Whirlwind Tour of Computational Geometry
by Ron Graham and Frances Yao 687

Minimal Surfaces Based on the Catenoid
by David Hoffman and William H. Meeks, III 702

Curvature in the Eighties
by Robert Osserman 731

Conway's Tiling Groups
by William P. Thurston 757

The cover illustration is a minimal surface discovered by David Hoffman (Amherst) and Hermann Karcher (Bonn). It is singly-periodic and behaves at infinity much like the union of helicoids. The image was made by James Riordan and Jim Hoffman (GANG, University of Massachusetts, Amherst).

THE MATHEMATICAL ASSOCIATION OF AMERICA

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 9 November 1990

Contents

(ISSN 0002-9890)

ARTICLES

Toetjes TOMAS FEDER 785

Set Isometries and Their Extensions in the Max Metric REGINA B. COHEN 795

On Stability of Invariant Subspaces of Matrices A. C. M. RAN AND L. RODMAN 809

LETTERS TO THE EDITOR 824

NOTES

A New Derivation of Stirling's Approximation of $n!$
GEORGE MARSAGLIA AND JOHN C. W. MARSAGLIA 826

Fibonacci Numbers and Aitken Sequences Revisited M. J. JAMIESON 829

A Cayley Theorem for Boolean Algebras
STEPHEN L. BLOOM, ZOLTÁN ÉSIK, AND ERNEST G. MANES 831

Rouché's Theorem for Polynomials MICHAEL FILASETA 834

THE TEACHING OF MATHEMATICS

A Simple Derivation of the Maclaurin Series
for Sine and Cosine DENG BO 836

An Infinite Product Expansion for the Square Root Function ERIC WINGLER 836

Some Models of Plane Geometries BRANKO GRÜNBAUM AND JAN MYCIELSKI 839

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 847

Advanced Problems and Solutions 857

REVIEWS

Learning Discrete Mathematics With ISETL.
By Nancy Baxter, Ed Dubinsky, and Gary Levin DONALD L. MEUNCH 861

Generatingfunctionology.
By Herbert S. Wilf RICHARD P. STANLEY 864

TELEGRAPHIC REVIEWS 867

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Macalester College, St. Paul, MN 55105. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

Toetjes

TOMÁS FEDER,* *Computer Science Department, Stanford University, Stanford, CA 94305*

TOMÁS FEDER received his B.S. in Mathematical and Computational Science from Stanford University, California. He is currently working towards his Ph.D. in Computer Science at Stanford University. He has also worked at the Xerox Palo Alto Research Center, the IBM Almaden Research Center in San Jose, and AT&T Bell Laboratories at Murray Hill. His research interests are in theoretical computer science, particularly algorithms and computational complexity.



Abstract. A number is secretly chosen from the interval [0, 1], and n players try in turn to guess this number. When the secret number is revealed, the player with the closest guess wins. We describe an optimal strategy for a version of this game.

1. Introduction. The Toetjes Problem was posed at the Programming and Problem Solving Seminar [1] held at Stanford University in the winter of 1987. It was suggested by Sape Mullender, who described it as follows:

> In Amsterdam, where I grew up, dessert is usually referred to as "toetje" (Dutch for "afters"). The problem of allocating a leftover toetje to one of the children in my family became the Toetjes Problem. The algorithm was the following: First my mother would choose a secret number between one and a hundred. Then the children, in turn, from youngest to oldest, would try to guess the number. After the last guess my mother would indicate whose guess was closest to her secret number and the winner would get the toetje.
>
> Now that I have a degree in mathematics, the problem still puzzles me: Given that the secret number is chosen randomly from the interval [0, 1], what is the optimal strategy for choosing a number for the ith child in a family of n children? The ith child knows what the first $i - 1$ children chose, and knows that all the children choose optimally (i.e., choose to maximize their own chance without consideration for the chances of any other child in particular).

We will study Mullender's "continuous" version of the problem, in which the secret number is chosen randomly from the interval [0, 1]. Notice that the description we have given so far is, in fact, incomplete. A first ambiguity arises because the optimal move for a player may not be unique. If this is the case, the criterion that the player uses to choose among equally desirable moves can influence the strategy followed by earlier players. We will resolve this ambiguity by assuming that there is a given *tie-breaking rule* that all players must follow when choosing among several equally desirable moves.

The choice of the tie-breaking rule affects the play quite drastically. To illustrate this, let $x_1, \ldots, x_n$ be the (distinct) numbers chosen by the players in the

*This research was primarily supported by a Bell Laboratories Scholarship, and in part by the National Science Foundation under grant CCR-86-10181, and by Office of Naval Research contract N00014-87-K-0502.

interval [0, 1]. Suppose player n plays between two earlier players i and j such that no other player has already played between them. Then his payoff (i.e., his probability of winning) is the same no matter where in the interval (x_i, x_j) he plays: it equals $(x_j - x_i)/2$, half the interval size. However, his choice will affect the payoff (and hence the strategy) for players i and j.

A second difficulty occurs because, in this continuous version of the problem, each player has an infinite number of choices, so there may not be an optimal move. Instead, it can happen that for each possible choice there is another choice that gives a slightly better payoff. For example, in the two-player game, suppose the first player chooses $1/2$; then the second player will get a better payoff the closer he plays to the first player, so there is no "optimal" move, but instead a family of moves arbitrarily close to optimal. We will later see that there are situations where a player can obtain a better payoff the closer he plays to a certain number ξ, while playing exactly at ξ would give him a very low payoff. We handle this difficulty by defining an optimal *limiting play* which is such that all players can obtain payoffs arbitrarily close to optimal by playing sufficiently close to this limiting play.

In this paper we characterize the optimal strategy under two simple tie-breaking rules: the "closest to the first player" rule and the "rightmost" rule. In particular, we prove that under the first of these tie-breaking rules, the payoff for the first and last players is exactly half the payoff for each of the remaining players. Under the second tie-breaking rule, the last two players get a payoff equal to half the payoff of each of the other players.

The remainder of the paper is organized as follows. In section 2 we give a rigorous definition of the game, which resolves the two ambiguities mentioned above. In section 3 we develop a plausible optimal playing strategy by making some assumptions about the tie-breaking rule. Then in section 4 we prove that the strategy of section 3 is in fact optimal under a simple tie-breaking rule, the "closest to the first player" rule. Section 5 extends the results of the previous section to the "rightmost" tie-breaking rule. Section 6 explores the "random" tie-breaking rule.

2. Definitions. We now give a precise definition of the "optimal" playing strategy. Given the k distinct numbers $x_1, \ldots, x_k$ chosen by the first k players, we consider *limiting plays* for the remaining $n - k$ players of the form $(x_{k+1}, \ldots, x_n, \prec)$, where $\prec$ is an ordering of the numbers $x_1, \ldots, x_n$ that extends the usual $<$ ordering into a total ordering. Intuitively, a limiting play of this form indicates that the last $n - k$ players play arbitrarily close to $x_{k+1}, \ldots, x_n$ and in such a way that the resulting ordering of the n players in the $[0, 1]$ interval is given by the total ordering $\prec$. While the numbers played are all distinct, some of the $n - k$ limiting values can in fact be equal to each other or to the numbers chosen by the first k players. In the case where two or more values are equal, the left-to-right order of the n moves in the $[0, 1]$ interval needs to be specified: only then can the payoff for each player corresponding to this limiting play be evaluated. This is the reason for including in the limiting play $(x_{k+1}, \ldots, x_n, \prec)$ an ordering $\prec$ that extends the partial ordering $<$ on the numbers $x_1, \ldots, x_n$ to a total ordering.

The winning probability (or payoff) $p(i, x_1, \ldots, x_n, \prec)$ for player i under such a limiting play can now be easily computed [see FIG. 1]: (a) If the move x_i is between x_j and x_k in the $\prec$ ordering, then the payoff is $(x_i - x_j)/2 + (x_k - x_i)/2 =$

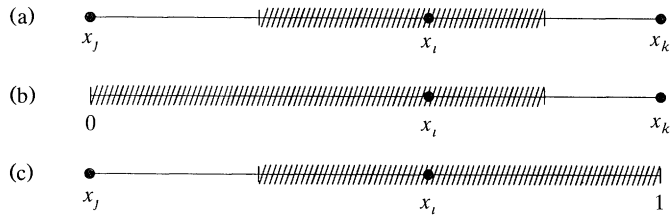


FIG. 1. Payoff for player i.

$(x_k - x_j)/2$. (b) If x_i is the smallest move and is followed by x_k, then the payoff is $(x_i - 0) + (x_k - x_i)/2 = (x_i + x_k)/2$. (c) If x_i is the largest move and is preceded by x_j, then the payoff is $(x_i - x_j)/2 + (1 - x_i) = 1 - (x_i + x_j)/2$. We use this payoff function to define the optimal limiting play $\text{opt}(x_1, \ldots, x_k) = (x_{k+1}, \ldots, x_n, \prec)$ for the last $n - k$ players, given the k distinct numbers $x_1, \ldots, x_k$ chosen by the first k players. The definition is by induction on the number $n - k$ of players left to play. The base case $k = n$ is simply $\text{opt}(x_1, \ldots, x_n) = (\prec)$, where $\prec$ is the $<$ ordering on the distinct values $x_1, \ldots, x_n$. Now we assume inductively that the function $\text{opt}(x_1, \ldots, x_k)$ has been defined for a given k, and we define $\text{opt}(x_1, \ldots, x_{k-1})$.

So fix the numbers $x_1, \ldots, x_{k-1}$ chosen by the first $k - 1$ players. We say that $(x_k, \ldots, x_n, \prec)$ is a *limiting play available to player k* if for all $\varepsilon > 0$ there are numbers $y_k, \ldots, y_n$ with $|y_i - x_i| < \varepsilon$ such that $\text{opt}(x_1, \ldots, x_{k-1}, y_k) = (y_{k+1}, \ldots, y_n, \prec)$. In other words, the limiting play is available to player k if player k can play close to x_k so that the remaining players will then play close to $x_{k+1}, \ldots, x_n$.

The optimal limiting payoff α for player k is now defined to be the least upper bound of the payoff $p(k, x_1, \ldots, x_n, \prec)$ over all limiting plays $(x_k, \ldots, x_n, \prec)$ available to player k. We can show that there exists a limiting play available to player k which achieves this maximum payoff α by considering a sequence of limiting plays available to player k that achieve payoffs arbitrarily close to α, and then using the Bolzano-Weierstrass theorem to obtain a limit point $(x_k, \ldots, x_n, \prec)$ for this sequence. We let S be the set of limiting plays $(x_k, \ldots, x_n, \prec)$ available to player k which give player k this maximum payoff α; thus S is the set of optimal limiting plays for player k. It is not difficult to show that S is in fact a closed set[1].

We can now apply any chosen *tie-breaking rule* to pick an optimal limiting play $(x_k, \ldots, x_n, \prec)$ from S. We could, for example, always pick a limiting play from S that has a largest x_k value (such a play exists because S is a closed set); this is in fact a particular case of the family of "rightmost" tie-breaking rules that we will consider in section 5. The chosen $(x_k, \ldots, x_n, \prec)$ is then the value of $\text{opt}(x_1, \ldots, x_{k-1})$, completing the inductive definition.

Finally, if all players play optimally and follow the given tie-breaking rule to choose among equally desirable moves, their numbers will be given in the limit by $\text{opt}(\,) = (x_1, \ldots, x_n, \prec)$, with corresponding winning probabilities $p(i, x_1, \ldots, x_n, \prec)$.

[1]That is, if for all $\varepsilon > 0$ there are y_i with $|y_i - x_i| < \varepsilon$ such that $(y_k, \ldots, y_n, \prec) \in S$, then $(x_k, \ldots, x_n, \prec) \in S$.

3. A candidate strategy. Finding a constructive characterization of the optimal strategy $\text{opt}(x_1, \ldots, x_k)$ is, under many tie-breaking rules, a difficult task. It turns out that for the "closest to the first player" tie-breaking rule considered in the next section the optimal strategy is, surprisingly, fairly simple. However, instead of just giving the strategy and proving that it is optimal (we will do this in the next section), we would like to describe here the properties of the tie-breaking rule which can lead us to guess such a strategy. These properties might in turn be useful for finding optimal strategies under other possible tie-breaking rules.

After examining the problem for small values of n, one observes that many tie-breaking rules seem to obey some form of the following *optimality principle*: *A player who does not play optimally for himself can only increase the chances of winning for the players that follow him.*

We use restricted versions of this principle to obtain a plausible playing strategy. When applied to the last player, this principle says that the chances for the last player are worst when the players before him play optimally (since his chances only increase when those players play suboptimally). We thus make the assumption that all players play so as to make the payoff of the last player as small as possible.

The optimal payoff for the last player given the numbers chosen by the players before him is easy to compute. Suppose the moves of the first $n-1$ players have partitioned $[0, 1]$ into n intervals of respective lengths $a_1, \ldots, a_n$, from left to right. Then the optimal limiting move for the last player gives him a payoff of $\max(a_1, a_2/2, \ldots, a_{n-1}/2, a_n)$, since the payoff when playing anywhere between two other players is half the distance between the two players, while the payoff when playing in the leftmost (rightmost) interval can be made arbitrarily close to the interval size by playing sufficiently close to the leftmost (resp., rightmost) player. Now suppose that the first $k-1$ players have already played, for some $k \geqslant 2$, partitioning $[0, 1]$ into k intervals $J_1, \ldots, J_k$ of respective sizes $c_1, \ldots, c_k$, from left to right. Given any real $B > 0$, the payoff of the last player can be forced to be $B/2$ or less by the $n-k$ players before him that are left to play, provided that at least $\lceil \rho_i \rceil - 1$ players move into the interval J_i for each $1 \leqslant i \leqslant k$. Here ρ_i is the ratio

$$\rho_i = \begin{cases} c_i/B & \text{if } 1 < i < k, \\ c_i/B + 1/2 & \text{if } i = 1 \text{ or } i = k. \end{cases} \tag{1}$$

This is so because in the case $1 < i < k$ the interval J_i must be partitioned into subintervals of size B or less, while in the cases $i = 1$ and $i = k$ the subintervals at both ends of $[0, 1]$ must be of size $B/2$ or less, to ensure that the payoff for the last player is at most $B/2$. This can be achieved if and only if

$$\sum_{1 \leqslant i \leqslant k} (\lceil \rho_i \rceil - 1) < r, \tag{2}$$

where $r = n - k + 1$ is the number of players left to play. If, following the optimality principle, the players before the last one play so as to make the payoff $\hat{B}/2$ of the last player as small as possible, then $\hat{B}$ must be the least value of B satisfying (2). An equivalent definition for this key parameter is that $\hat{B}$ is the

largest B satisfying

$$\sum_{1 \leqslant i \leqslant k} \lfloor \rho_i \rfloor \geqslant r; \tag{3}$$

in fact $\hat{B}$ is the only value of B which satisfies (2) and (3) simultaneously. This is because the summands in (2) and (3) are equal unless ρ_i is an integer, in which case they differ by 1. Since only r players are left to play, it is natural in view of (3) to allocate at most $\lfloor \rho_i \rfloor$ players to the interval J_i, including player n. Then if r_i out of the last r players move into the interval J_i, we will have

$$\lceil \rho_i \rceil - 1 \leqslant r_i \leqslant \lfloor \rho_i \rfloor \quad \text{for all } 1 \leqslant i \leqslant k, \quad \text{with} \sum_{1 \leqslant i \leqslant k} r_i = r. \tag{4}$$

We will try to determine more precisely where these r_i players should play inside the interval J_i, for $1 < i < k$. The first $r_i - 1$ players to play in J_i partition this interval into r_i subintervals. There must be a subinterval J_i' among these of size at least $\hat{B}$, since $r_i \leqslant \lfloor c_i/\hat{B} \rfloor$ by (4). The last player p to play in J_i should then be able to obtain a payoff of at least $\hat{B}/2$ by playing inside J_i'. Now the optimality principle applied to player p says that the $r_i - 1$ players that play in the interval J_i before player p will play so as to make the size of the subinterval J_i' as small as possible. However, the other subintervals of J_i should not be made larger than $\hat{B}$ because otherwise player n could move into J_i and obtain a payoff greater than $\hat{B}/2$. Thus the first $r_i - 1$ players in J_i should partition this interval into $r_i - 1$ subintervals of size $\hat{B}$, and one subinterval J_i' of size at least $\hat{B}$ for player p. If we assume that player p plays in J_i' so as to create a subinterval of size $\hat{B}$ as well, we obtain a simple playing strategy: *The r_i players that play in $J_i, 1 < i < k$, partition this interval into r_i subintervals of size $\hat{B}$ and a leftover subinterval K_i of size at most $\hat{B}$.*

This is illustrated in FIG. 2. A subinterval of size $\hat{B}/2$ (instead of $\hat{B}$) is used at both ends of [0, 1] to keep the potential payoff for the last player bounded by $\hat{B}/2$. Also notice that we have placed the leftover subinterval K_i in each J_i as close as possible to the number x_1 played by player 1; this will be justified by the tie-breaking rule of the next section.

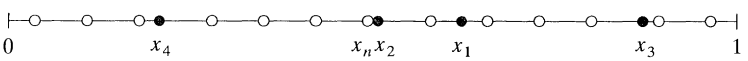


FIG. 2. Optimal play $(x_k, \ldots, x_n, \prec)$ given $(x_1, \ldots, x_{k-1})$ for $n = 17$ and $k = 5$, under the "closest to the first player" tie-breaking rule. Here $x_n = x_2$ with $x_n \prec x_2$.

4. Optimal strategy under the "closest to the first player" tie-breaking rule. In this section we characterize the optimal playing strategy under the "closest to the first player" tie-breaking rule. This tie-breaking rule assumes that the optimal limiting play $(x_k, \ldots, x_n, \prec)$ chosen by a player $k \geqslant 2$ is such that no other optimal limiting play that has the first k players in the same relative order in the interval [0, 1] as the chosen play can have player k closer to the first player. Notice that we make no assumptions about how the tie-breaking rule chooses among equivalent moves for the first player, or about how it chooses the relative position of a player $k \geqslant 2$ with respect to earlier players if several optimal alternatives are available: It turns out that the optimal strategy is essentially the same regardless of

how these choices are made. Hence our solution applies not just to a single tie-breaking rule, but rather to a family of possible tie-breaking rules.

We recall some definitions from the previous section. Suppose for now that $k \geqslant 2$. The numbers $x_1, \ldots, x_{k-1}$ chosen by the first $k-1$ players partition $[0, 1]$ into k intervals $J_1, \ldots, J_k$, from left to right, of respective lengths $c_1, \ldots, c_k$. Let $\hat{B}$ be the unique value of B satisfying both (2) and (3), where the ρ_i are given as functions of B by (1), and $r = n - k + 1$; then $\hat{B}$ is also the smallest B satisfying (2) and the largest B satisfying (3).

We define $\lfloor \rho_i \rfloor$ *marked positions* $J_i(1), \ldots, J_i(\lfloor \rho_i \rfloor)$ inside each J_i. Let $J_i(0)$ and $J_i(\lfloor \rho_i \rfloor + 1)$ be the endpoints of J_i which are farthest from and closest to x_1, respectively. Then the marked positions inside J_i are given, for $j = 1, 2, \ldots, \lfloor \rho_i \rfloor$, by (see FIG. 2)

$$|J_i(j) - J_i(0)| = \begin{cases} j\hat{B} & \text{if } 1 < i < k, \\ (j - 1/2)\hat{B} & \text{if } i = 1 \text{ or } i = k. \end{cases} \tag{5}$$

We associate with each marked position $J_i(j)$ the payoff that a player at $J_i(j)$ would get if each marked position was played by one player. Therefore, the payoff of every marked position equals $\hat{B}$, except for the last marked position $J_i(\lfloor \rho_i \rfloor)$ of each interval J_i; this last position has a payoff of $(\hat{B} + l(i))/2$, where $l(i)$ is the size of the leftover interval $K_i = (J_i(\lfloor \rho_i \rfloor), J_i(\lfloor \rho_i \rfloor + 1))$ and is given by

$$l(i) = \begin{cases} c_i \bmod \hat{B} & \text{if } 1 < i < k, \\ (c_i + \hat{B}/2) \bmod \hat{B} & \text{if } i = 1 \text{ or } i = k. \end{cases} \tag{6}$$

We can now state our main result:

THEOREM 1. *The optimal limiting play under the "closest to the first player" tie-breaking rule is given for* $k \geqslant 2$ *by* $\operatorname{opt}(x_1, \ldots, x_{k-1}) = (x_k, \ldots, x_n, \prec)$, *where* $x_k, \ldots, x_n$ *are* $n - k + 1$ *marked positions of highest associated payoff, in order of decreasing payoff.*

Proof. Let us first determine which marked positions will be played according to this theorem. Notice that the total number of marked positions is at least r, the number of players left to play, by equation (3). Among these marked positions, those of least payoff are the positions $J_i(\lfloor \rho_i \rfloor)$ with ρ_i an integer, which have $l(i) = 0$ and a payoff of $\hat{B}/2$. The other marked positions will thus be used first; however, there are only $\lceil \rho_i \rceil - 1$ of these other marked positions in each J_i, and hence less than r in total by (2), so the theorem implies that one or more of the marked positions of low payoff $\hat{B}/2$ will have to be used as well. From these remarks, it follows that the number of players that will end up in each J_i (according to the theorem) satisfies (4).

We will prove the theorem by induction on the number $n - k + 1$ of players left to play. For the base case $k = n$, the marked positions are endpoints $J_i(\lfloor \rho_i \rfloor + 1)$ of highest payoff, and these are also the optimal moves for the last player under the "closest to the first player" tie-breaking rule. For the inductive step, suppose without loss of generality that player $k < n$ plays in an interval J_i with the endpoint $J_i(0)$ to the left and the endpoint $J_i(\lfloor \rho_i \rfloor + 1)$ to the right. Then J_i is split into a left interval $L = (J_i(0), x_k)$ and a right interval $R = (x_k, J_i(\lfloor \rho_i \rfloor + 1))$. The

new set of $k + 1$ intervals will then have a new parameter $\hat{B}'$ satisfying (2) and (3), and we can assume inductively that the last $n - k$ players will play as indicated by the theorem for this new parameter $\hat{B}'$. Notice that $\hat{B}' \geqslant \hat{B}$, since $\hat{B}$ is still a solution to (3) for this new set of intervals, while $\hat{B}'$ is by definition the largest solution to (3).

LEMMA 1 (Low payoff for unmarked positions). *If player k plays between $J_i(j-1)$ and $J_i(j)$ (or at $J_i(j)$), his payoff will be no greater than the payoff associated with marked position $J_i(j)$, for $1 \leqslant j \leqslant \lfloor \rho_i \rfloor$, and smaller than $\hat{B}/2$ for $j = \lfloor \rho_i \rfloor + 1$.*

Proof. We claim that if player k plays at x_k, between $J_i(j-1)$ and $J_i(j)$ (or at $J_i(j)$), then for $j > 1$ some player will play in L at $J_i(j-1)$ or higher, and for $j < \lfloor \rho_i \rfloor$ some player will play in R at $J_i(j+1)$ or lower. This claim implies the lemma: For $1 \leqslant j \leqslant \lfloor \rho_i \rfloor$, the payoff associated with marked position $J_i(j)$ is at least as high, since this payoff corresponds to a situation where $x_k = J_i(j)$ and no player plays between $J_i(j-1)$ and $J_i(j+1)$. For $j = \lfloor \rho_i \rfloor + 1$, either $j > 1$, in which case the payoff is less than $\hat{B}/2$ because $l(i) < \hat{B}$, or $j = 1$, in which case the payoff is less than $\hat{B}/2$ because $\lfloor \rho_i \rfloor = 0$ so that J_i is a small interval.

To prove the claim, we consider two cases. If $\hat{B}' = \hat{B}$, then some player after player k will play at $J_i(j-1)$, for $j > 1$. Also, some player will play at distance $\hat{B}$ from x_k in R, and hence at $J_i(j+1)$ or lower, for $j < \lfloor \rho_i \rfloor$. The remaining case is $\hat{B}' > \hat{B}$. Since the new parameter $\hat{B}'$ is greater than $\hat{B}$, the number of players that can play in each interval is now at most $j - 1$ for L, at most $\lceil \rho_i \rceil - j$ for R, and at most $\lceil \rho_{i'} \rceil - 1$ for each of the other intervals $J_{i'}$ with $i' \neq i$. By equation (2), these numbers can only add up to $r - 1$, the number of players left to play after player k, if exactly $j - 1$ players play in L, exactly $\lceil \rho_i \rceil - j$ in R, and exactly $\lceil \rho_{i'} \rceil - 1$ in the remaining $J_{i'}$. Then, for $j > 1$, the $(j-1)$th player in L (from left to right) is above $J_i(j-1)$. Also, for $j < \lfloor \rho_i \rfloor$, if the leftmost player in R were at $J_i(j+1)$ or above, then there could be no more than $\lceil \rho_i \rceil - j - 1$ players in R, since $\hat{B}' > \hat{B}$, so the leftmost player in R must in fact be below $J_i(j+1)$. This proves the claim and completes the proof of the lemma. □

LEMMA 2 (Preservation of marked positions). *Let $J_i(j)$ be a marked position of highest associated payoff. Then among the limiting moves $(x_k, \ldots, x_n, \prec)$ available to player k there is one in which $x_k = J_i(j)$, and $x_k, \ldots, x_n$ are $n - k + 1$ marked positions of highest associated payoff, in order of decreasing payoff.*

Proof. If player k plays at a marked position $J_i(j)$ (this is feasible unless $l(i) = 0$), then $\hat{B}$ also satisfies (2) for the new set of $k + 1$ intervals, so $\hat{B}' = \hat{B}$. The new set of $k + 1$ intervals therefore has a new marked position $L(j)$ at $J_i(j)$, with $L(j) \prec J_i(j)$. If none of the remaining players plays at $L(j)$, then they will all play at the old marked positions in order of decreasing payoff, as desired. On the other hand, if one of them plays at $L(j)$, then the condition of the lemma is not satisfied and in fact player k obtains a low payoff of $\hat{B}/2$ or less.

However, it is still possible for player k to obtain the limiting play described in the lemma by playing just below $J_i(j)$, at $J_i(j) - \varepsilon$, where $\varepsilon < \hat{B} - l(i)$. Now if $\hat{B}' = \hat{B}$, then the marked positions remain the same after player k has played, except for the marked positions $J_i(j')$ with $j' > j$, which change to $R(j' - j) = J_i(j') - \varepsilon$. Letting ε go to zero gives a limiting play as stated in the lemma. The

remaining case is $\hat{B}' > \hat{B}$. Then, as in the proof of Lemma 1, we must have $j - 1$ players in L, $\lceil \rho_i \rceil - j$ players in R, and $\lceil \rho_{i'} \rceil - 1$ players in the remaining $J_{i'}$. As before, the leftmost player in R must be below $J_i(j+1)$ (provided that $\lceil \rho_i \rceil - j > 0$ so that R has at least one player), so $\hat{B}' < \hat{B} + \varepsilon$; if we let ε go to zero then $\hat{B}'$ approaches $\hat{B}$ and the positions of the players approach the marked positions of highest payoff. If $\lceil \rho_i \rceil - j = 0$, then $J_i(j)$ is a marked position with payoff $\hat{B}/2$ and can only be a position of highest payoff if $\lceil \rho_{i'} \rceil - 1$ equals zero for all intervals $J_{i'}$. But $\lceil \rho_{i'} \rceil - 1$ is the number of players in each $J_{i'}$, so player k would be the last player, contrary to the assumption that $k < n$. □

We use these two lemmas to prove the theorem. If α is the payoff associated with a marked position $J_i(j)$ of highest payoff, then player k can obtain a payoff of α in the limit by playing at $J_i(j)$ or just below $J_i(j)$ as in Lemma 2. On the other hand, by Lemma 1, the only plays that could possibly give the same payoff α to player k involve playing between $J_i(j-1)$ and $J_i(j)$, where $J_i(j)$ is a marked position of payoff α, but then position $J_i(j)$ is closer to x_1 and hence should be preferred by player k under the "closest to the first player" tie-breaking rule. Thus the optimal limiting play is the one given in Lemma 2, completing the induction and proving the theorem. □

To complete the analysis, we must examine the optimal strategy for the first player. Consider the $n - 1$ positions $p_i = (2i-1)/2(n-1)$ for $1 \leqslant i \leqslant n-1$. The above theorem implies that if player 1 plays at one of these positions, then the remaining $n - 1$ players will play in the limit at these $n - 1$ positions (one of them coinciding with player 1), giving player 1 a payoff of $1/2(n-1)$ [see FIG. 3]. If on the other hand player 1 does not play at one of these positions, the remaining $n - 1$ players will play near these $n - 1$ positions but closer to player 1, giving him a payoff smaller than $1/2(n-1)$. This gives:

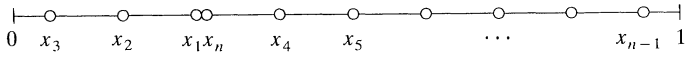


FIG. 3. Optimal play $(x_1, \ldots, x_n, \prec)$ for $n = 10$ under the "closest to the first player" tie-breaking rule. Here $x_n = x_1$ with $x_1 \prec x_n$.

COROLLARY 1. *In the optimal limiting play* $\text{opt}(\,) = (x_1, \ldots, x_n, \prec)$ *under the "closest to the first player" tie-breaking rule, the limiting moves* $x_1, \ldots, x_{n-1}$ *of the first* $n - 1$ *players are the numbers* $1/2(n-1), 3/2(n-1), \ldots, (2n-3)/2(n-1)$, *in some order, and the limiting move* x_n *of the last player equals* x_1. *This gives a limiting payoff of* $1/2(n-1)$ *for the first and last players, and* $1/(n-1)$ *for the remaining players.*

Thus the youngest and the oldest children will be less likely to get the leftover toetje.

The preceding analysis can in fact be carried through even if the choice among equally desirable intervals is made at random or nondeterministically (i.e., chosen arbitrarily by each player when it is his turn to play), and even if only the last players—those that are left to play once each interval has at most one marked position—use the "closest to the first player" tie-breaking rule to select a position inside the chosen interval.

5. The "rightmost" tie-breaking rule. We now consider the "rightmost" tie-breaking rule. This tie-breaking rule assumes that the optimal limiting play $(x_k, \ldots, x_n, \prec)$ chosen by a player k is such that no other optimal limiting play that has the first k players in the same relative order in the interval $[0, 1]$ as the chosen play can have player k further to the right. Again, we make no assumptions about how the tie-breaking rule chooses the relative position of a player with respect to earlier players if several optimal alternatives are available, so our solution actually applies to a family of tie-breaking rules.

It turns out that the solution for the "rightmost" tie-breaking rule is very similar to the solution for the "closest to the first player" rule, provided that we make a few changes to the marked positions. We again consider the intervals $J_1, \ldots, J_k$, and define marked positions as follows. Let $J_i(0)$ and $J_i(\lfloor \rho_i \rfloor + 1)$ be the left and the right endpoints of J_i respectively. If $i < k$, the $\lfloor \rho_i \rfloor$ marked positions for J_i and their associated payoffs are defined as before. For J_k, on the other hand, we define $\lfloor \rho_k \rfloor$ marked positions by letting $J_k(j) = J_k(0) + j\hat{B}$ if $1 \leqslant j < \lfloor \rho_k \rfloor$ and $J_k(\lfloor \rho_i \rfloor) = J_k(\lfloor \rho_i \rfloor + 1) - \hat{B}/2$. Therefore, the associated payoff of every marked position in J_k equals $\hat{B}$, except for the last two marked positions $J_k(\lfloor \rho_k \rfloor - 1)$ and $J_k(\lfloor \rho_k \rfloor)$; these two positions have a payoff of $(\hat{B} + l(k))/2$, where $l(k)$ is the size of the leftover interval $K_i = (J_k(\lfloor \rho_k \rfloor - 1), J_k(\lfloor \rho_k \rfloor))$ and is given by $l(k) = (c_k + \hat{B}/2) \bmod \hat{B}$. In particular, if ρ_k is an integer with $\rho_k \geqslant 2$, then we have a *double marked position* $J_k(\lfloor \rho_k \rfloor - 1) = J_k(\lfloor \rho_k \rfloor)$ with $J_k(\lfloor \rho_k \rfloor - 1) \prec J_k(\lfloor \rho_k \rfloor)$, and each of these two positions has an associated payoff of $\hat{B}/2$. For the particular case $\rho_k = 2$, we will also need to consider a *special position* $J_k(\,)$ just to the right of $J_k(0)$, that is, $J_k(\,) = J_k(0)$ with $J_k(0) \prec J_k(\,)$; this special position is not considered a marked position.

Using this new set of marked positions, the results of the previous section for the "closest to the first player" tie-breaking rule can be extended to the "rightmost" tie-breaking rule as well, provided that the somewhat subtle situations that arise when a double marked position is present are properly treated.

THEOREM 2. *The optimal limiting play under the "rightmost" tie-breaking rule is given for* $k \geqslant 2$ *by* $\text{opt}(x_1, \ldots, x_{k-1}) = (x_k, \ldots, x_n, \prec)$, *where one of the following conditions holds*:

1. *There is no double marked position, and* $x_k, \ldots, x_n$ *are marked positions of highest payoff, in order of decreasing payoff.*
2. *There is a double marked position, two* x_i *among* $x_k, \ldots, x_n$ *are at the double marked position, and* $x_k, \ldots, x_n$ *are marked positions of highest payoff, in order of decreasing payoff.*
3. *There is a double marked position, only one* x_i *among* $x_k, \ldots, x_{n-1}$ *is at the double marked position, and* $x_k, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ *are marked positions of highest payoff, in order of decreasing payoff.*
4. *There is a special position,* x_n *is at the special position, no* x_i *is at the double marked position, and* $x_k, \ldots, x_{n-1}$ *are marked positions of highest payoff, in order of decreasing payoff.*

If we now consider, for each possible move of the first player, the optimal play for the remaining players described by the four cases in the theorem, we obtain:

COROLLARY 2. *In the optimal limiting play* $\text{opt}(\,) = (x_1, \ldots, x_n, \prec)$ *under the "rightmost" tie-breaking rule, the limiting moves* $x_1, \ldots, x_{n-1}$ *of the first* $n - 1$

players are the numbers $1/2(n-1), 3/2(n-1), \ldots, (2n-3)/2(n-1)$, *in some order, and the limiting move* x_n *of the last player equals* x_{n-1}. *This gives a limiting payoff of* $1/2(n-1)$ *for the last two players, and* $1/(n-1)$ *for the remaining players.*

6. Other tie-breaking rules. For other tie-breaking rules, the situation is more complicated. To simplify things and avoid dealing with the special intervals J_1 and J_k, let us assume that for all the intervals J_i, both endpoints of J_i have been taken by earlier players. This will be the case if the numbers 0 and 1 are taken at the beginning of the game (or, alternatively, if the game is played on a circle).

Under the simplifying assumption that the endpoints have been taken at the beginning of the game, there is no difference between the "closest to the first player" and the "rightmost" tie-breaking rules. We now examine the situation under the more equitable tie-breaking rule in which the choice among equivalent moves is made uniformly at random.

If we follow the heuristic argument of section 3, we would expect that the number of players r_i in each J_i would still satisfy (4). Under the "random" tie-breaking rule, however, the subintervals of size $\hat{B}$ would be taken by the first $r_i - 1$ players in J_i at random from either the left end or the right end of J_i, leaving a subinterval J_i' of size greater than $\hat{B}$ somewhere in between. Then the last player in J_i would play at random inside J_i' but in such a way as to partition J_i' into two subintervals of size at most $\hat{B}$.

This is indeed the optimal strategy when the total number of players in each interval is small. The inductive step, however, fails when the number of players is large. Thus, in the seven-player game with both endpoints taken, if all players follow the strategy we described, the expected payoff for the first player will be $(1/7)(1 - (1/2)^7) < 1/7$, but if the first player chooses the midpoint of the $[0, 1]$ interval and the remaining six players follow the strategy we described, playing three to each side of the first player, the expected payoff for the first player will be $(1/6)(1 - (1/2)^3) > 1/7$ (see [**1**] for the details of these calculations).

Nevertheless, since this strategy is optimal when the number of players is small, it may provide a good base case from which to extend the solution for the "random" tie-breaking rule to larger values of n.

7. Conclusion. We have shown that a simple zero-sum game in which all players play independently so as to maximize their own payoff can be viewed, under certain conditions, as a game in which the players cooperate to minimize the payoff of the last player. Furthermore, this approach is helpful in developing an optimal strategy for the game. An interesting question is whether this principle can be applied to a broader family of zero-sum games. A candidate problem for this approach may be the two-dimensional version of toetjes.

Acknowledgments. The author would like to thank Donald Knuth, Tomas Rokicki, and the other students in the problem seminar for their ideas and suggestions on this problem.

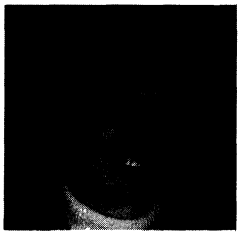
REFERENCES

1. Tomas G. Rokicki and Donald E. Knuth, A programming and problem-solving seminar, Stanford University Technical Report STAN-CS-87-1154 (1987).
2. Maxwell Carver, Three wise men and a rope, *Discover*, December 1988.

Set Isometries and Their Extensions in the Max Metric

REGINA B. COHEN, *Villanova University, Villanova, PA 19085*

REGINA B. COHEN: I received an A.B. from Goucher College and an M.A. from the University of Pennsylvania. After being out of mathematics for several years while raising a family, I began teaching at Villanova University in 1982. At present I am also a graduate student at Bryn Mawr College working on a problem in ergodic theory with Kyewon Park. The idea of a set isometry was suggested by the late S. M. Ulam in the context of mappings for paradoxical decompositions.



Introduction. The max and the taxicab metrics in $\mathbf{R}^2$ are accessible yet interesting examples of non-Euclidean metrics. Each is a specific case of the more general Minkowski metric where the defining unit "circle" is a centrally symmetric closed convex curve. It has been shown in [3] for the max metric ρ in $\mathbf{R}^n$ and in [6] for the taxicab metric d_T in $\mathbf{R}^2$ that the respective isometries of the whole space form a subgroup of the Euclidean isometries of the whole space. There are, nevertheless, many isometries from one set to another which cannot be extended to an isometry of the space onto itself and are evidently quite different from Euclidean isometries.

Let ρ be the metric in $\mathbf{R}^n$ given by $\rho(p, q) = \max_{1 \leqslant i \leqslant n} |x_i - y_i|$ where $p = (x_1, \ldots, x_n)$ and $q = (y_1, \ldots, y_n)$. For $E, F \subseteq \mathbf{R}^n$, $f\colon E \to F$ is a ρ-isometry and E and F are ρ-isometric or ρ-congruent if f is surjective and $\rho(f(p), f(q)) = \rho(p, q)$ for all points p and q in E. The usual Euclidean distance between p and q will be denoted by $e(p, q)$. f is a distance preserving bijection from one set to another set.

For the most part, this paper deals with the case of the max norm or metric in $\mathbf{R}^2$, but many results pertain to the max norm in $\mathbf{R}^n$. In addition, with methods which are natural extensions of those used in this paper, the reader will see that some proofs can be modified for a Minkowski metric where the defining unit "circle" is a polygon.

There is a class of simple $2n$-point sets in $\mathbf{R}^n$ (called γ-stars) on which a ρ-isometry must be Euclidean and can be extended uniquely to an unbounded set. Any $E \subseteq \mathbf{R}^n$ of positive Lebesgue measure contains many such sets. Then $F \subseteq \mathbf{R}^n$ whose closure has positive Lebesgue measure can be partitioned up to a set of measure zero into disjoint subsets, on each of which a ρ-isometry is Euclidean. Depending on the geometry and not the cardinality of the set involved, some ρ-isometries cannot be extended at all, others can be extended but not uniquely, still others can be extended uniquely to $\mathbf{R}^n$. Finally, another type of four-point set is introduced in $\mathbf{R}^2$; together with γ-stars they are minimal sets in $\mathbf{R}^2$ on which ρ-isometries must be Euclidean, answering a conjecture of [3] in the negative.

2. Examples in $\mathbf{R}^2$ and preliminaries.

EXAMPLE 1. Let $E = [-1/2, 1/2] \times \{0\}$ and $f((x, 0)) = (x, x^2)$. This is a ρ-isometry on E since, for $x \geqslant -1/2$ and $z \leqslant 1/2$, $|x + z| \leqslant 1$. Then $\rho((x, x^2), (z, z^2)) = \max(|x - z|, |x^2 - z^2|) = |x - z| = \rho((x, 0), (z, 0))$. So a straight line segment is ρ-congruent to a segment of a curve.

EXAMPLE 2. Let E be the x-axis and $0 < |m| \leqslant 1$. g: $(x, 0) \to (x, m|x|)$ and f: $(x, 0) \to (x, mx)$ are both ρ-isometries. Both E and $f(E)$ are vector subspaces, but it is easily seen that f cannot be extended by even one point: a point $p \notin E$ has infinitely many points on the x-axis at a distance a from p, whereas any point q not on the line $f(E)$ has at most two points on $f(E)$ at a distance a from q. Thus no point q could be the image of p under an extension of f. (See FIG. 1)

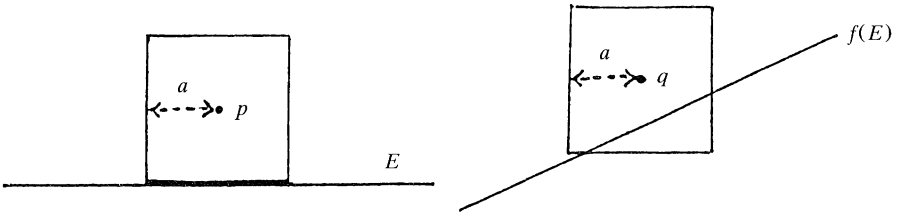


FIG. 1. A ρ-isometry which cannot be extended.

Analogously, let F be the y-axis and h: $(0, y) \to (my, y)$ for $0 < |m| \leqslant 1$. h is a ρ-isometry mapping the y-axis onto the line $y = (1/m)x$.

Remark. Since the parallelogram law does not hold, $(\mathbf{R}^2, \rho)$ is not an inner product space. The Witt Extension Theorem [**5**] states: If a metric vector space V is equipped with a non-degenerate quadratic form, then any isometry of a subspace U_1 onto a subspace U_2 can be extended to an isometry of $V \to V$. The term isometry here means a bijective linear map which preserves the quadratic form. The above example is an easily available illustration that such an extension theorem is not always possible in a vector space where the metric is given by a norm and not by an inner product.

EXAMPLE 3. $p_1 = (1, 1)$, $p_2 = (-1, 1)$, $p_3 = (-1, -1)$, $p_4 = (1, -1)$. $\rho(p_i, p_j) = 2$ for $i, j = 1, 2, 3, 4$. Any permutation f of the four points is a ρ-isometry; in particular, take $f(p_1) = p_4$, $f(p_2) = p_2$, $f(p_3) = p_3$ and $f(p_4) = p_1$.

EXAMPLE 4. The above ρ-isometry can be considered as a localization of the ρ-isometry g on $E = \{(x, y) | |x| \geqslant |y|\}$ where $g((x, y)) = (x, y)$ if $x \leqslant 0$ and $g((x, y)) = (x, -y)$ if $x > 0$. There is also the ρ-isometry h on the set $F = \{(x, y) | |y| \geqslant |x|\}$ where $h((x, y)) = (x, y)$ if $y \leqslant 0$ and $h(x, y) = (-x, y)$ if $y > 0$. The upper cone is reflected about the y-axis, the lower cone remains fixed.

A geometric feature of this metric gives rise to some of these unusual congruences. In $\mathbf{R}^2$ with the usual metric, every point can be uniquely located by specifying its distances from three non-collinear points. In the max metric this is not the case, nor is it so for any Minkowski metric where the defining unit "circle" contains a straight line segment. In FIG. 2(i), every point on the heavy line segment has distance a from p, b from q and c from r. Using this property of the max ρ or taxicab metric d_T, many examples can be constructed of three (or more) point sets, ([**4**], [**6**], and FIG. 2(ii)) which are ρ- or d_T-congruent as sets, but are not e-congruent. We shall see that it is true in general that for any three-point set in the plane there is a ρ-isometry which is not Euclidean. Since ρ-isometries of $\mathbf{R}^2 \to \mathbf{R}^2$ are Euclidean, there is no configuration of three points such that any ρ-isometry

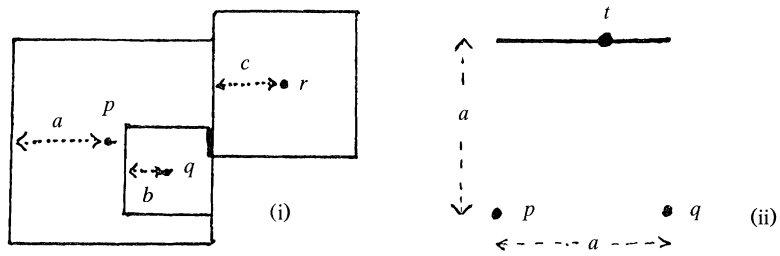


FIG. 2. (i) Points are not uniquely located by specifying distances from three noncollinear points. (ii) All sets with p, q fixed and t on the heavy line segment are ρ-congruent.

on those three points can be extended to a ρ-isometry of the space onto itself. We shall see later that, although ρ-isometries on some sets must be Euclidean, there is always a non-Euclidean extension of a ρ-isometry which is a relativization of an element from the group of symmetries of a square on a bounded set.

Some terminology will clarify the proofs of these assertions: given a point $p \in \mathbf{R}^2$, construct lines L and L' of slopes ± 1 through p, obtaining the four open quadrants $Q_r(p), Q_l(p), Q_u(p), Q_d(p)$ (right, left, up, down) of p. Note that if $p \in \overline{Q_r(q)}$ then $q \in \overline{Q_l(p)}$ (where the bar denotes closure with the usual topology) and the ρ-distance of p to q is the absolute value of the difference in their x- or horizontal coordinates. Also if $p \in \overline{Q_u(q)}$, then $q \in \overline{Q_d(p)}$, and their ρ-distance is the same as their vertical e-distance. Whenever $p \in \overline{Q_i(q)}$, $\overline{Q_i(p)} \subset \overline{Q_i(q)}$ for $i = r, l, u, d$.

Generalizing Example 4, a half-pivot about a point p is defined as one of four mappings with respect to the quadrants of p. For a right (left) half-pivot about p, the closed right (left) quadrant of p is reflected about a horizontal line through p while the closed left (right) quadrant stays fixed. Likewise, for a lower (upper) half pivot about p, the closed lower (upper) quadrant of p is reflected about a vertical line through p, while the closed upper (lower) quadrant stays fixed. Relativizations of these mappings are ρ-isometries for any sets contained in the mentioned quadrants: e.g., if $E \subseteq \overline{Q_u(p)} \cup \overline{Q_l(p)}$, then an upper or lower pivot about p is a ρ-isometry on E. See FIG. 3.

PROPOSITION 1. *For any three points in* $\mathbf{R}^2$ *there is a* ρ*-isometry which is not Euclidean.*

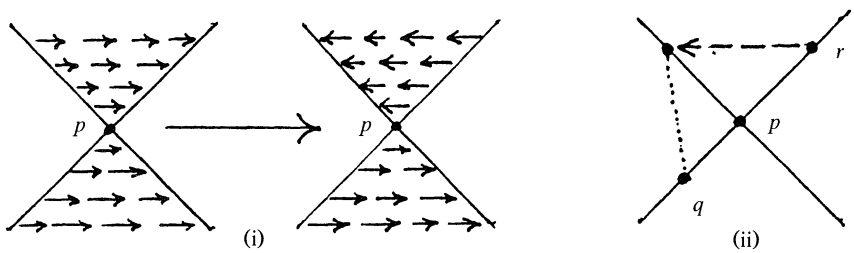


FIG. 3. (i) An upper half-pivot about p. (ii) $e(q, r)$ is equal to the sum of the e-lengths of the legs of the triangle pqr', and $e(q, r')$ is the hypotenuse, while $\rho(q, r) = \rho(q, r')$.

Case (a). If the three points p_1, p_2 and p_3 are collinear, map them into another line using the restriction of the ρ-isometry f, f^{-1}, h, or h^{-1} of Example 2 to the three points.

Case (b). Suppose two of the points, say p_1 and p_2, are on a line L of slope $+1$ or -1. Construct a line L' through p_3 perpendicular to L. $L \cap L' = p$, which could be coincident with p_1 or p_2. In any case, one of the two points is in one closed quadrant of p and the other two are in the opposite closed quadrant. Half-pivot about p so that p_3 is now on L, and either all three points are on L or p_1 is on L while p_2 is on L'. In either case, two points which had for e-distance the hypotenuse of a right triangle now have e-distance equal to the sum of the lengths of the legs.

Case (c). If no pair is on the same line of slope $+1$ or -1, then construct lines of slope $+1$ through all three points. One of the lines L_2 lies between the other two, L_1 and L_3. Label the points so that $p_1 \in L_1$, $p_2 \in L_2$, and $p_3 \in L_3$. Construct L_1' and L_3' through p_1 and p_3, respectively, with slope -1. $q_3 = L_2 \cap L_3'$ and $q_1 = L_2 \cap L_1'$. If p_2 is interior to the rectangle formed by L_1, L_3, L_1' and L_3', then a half-pivot about q_3 or a half-pivot about q_1 will change Euclidean distances, each bringing p_3 or p_1, respectively, onto L_2. If p_2 is not interior to the rectangle, choose whichever of q_1 or q_3 on L_2 is further from p_2. A right or left half-pivot about that point will bring p_3 or p_1 onto L_2, leaving p_2 fixed on L_2. ρ-distances are maintained (FIG. 4).

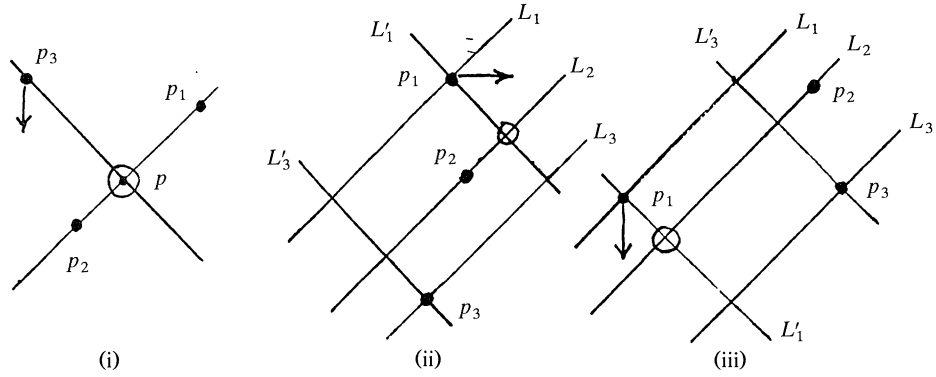


FIG. 4. (i) case (b). (ii) case (d). (iii) case (d). Half-pivot points are circled.

3. ρ-isometries on γ-stars and extensions. It is not difficult to find four-point sets on which ρ-isometries are not Euclidean (Example 3), but there are four-point sets on which ρ-isometries must be Euclidean. We shall see that configuration matters. Define a γ-star at $p = (x, y)$ to be the four-point set $(x + \gamma, y) = p_1$, $(x - \gamma, y) = p_3, (x, y + \gamma) = p_2$ and $(x, y - \gamma) = p_4$, where $\gamma > 0$. There is a simpler proof of Lemma 1 given in the proof of Prop. 4, but the method here is used later in the proof of Prop. 3.

LEMMA 1. *Let S be a γ-star at (x, y) and f a ρ-isometry on S. Then f is Euclidean, and $f(S)$ is a γ-star at p', the intersection of the line segment between $f(p_1)$ and $f(p_3)$*

and that between $f(p_2)$ and $f(p_4)$. f is a translation followed by an element from the group of the square.

Proof.

$$\rho(p_1, p_3) = \rho(p_2, p_4) = 2\gamma$$

and

$$\rho(p_1, p_2) = \rho(p_2, p_3) = \rho(p_3, p_4) = \rho(p_4, p_1) = \gamma \tag{1}$$

All squares mentioned have sides parallel to the axes. $f(p_1) = (x', y')$ can be considered a translation of p_1. $f(p_3)$ must be on a square S_2 of side 4γ centered at $f(p_1)$; suppose it is on the top side. Both $f(p_2)$ and $f(p_4)$ must be on a square S_1 side 2γ centered also at $f(p_1)$; since they are 2γ apart, however, they must be on opposite faces, right and left, of this square. $f(p_2)$ and $f(p_4)$ are at distance γ from $f(p_3)$, and must also be on opposite faces of a square S_3 of side 2γ centered at $f(p_3)$. The only configuration which will satisfy all the distance requirements is the γ-star centered at $(x', y' + \gamma)$: $f(p_1) = (x', y')$, $f(p_2) = (x' + \gamma, y' + \gamma)$, $f(p_3) = (x', y' + 2\gamma)$, $f(p_4) = (x' - \gamma, y' + \gamma)$. $f(p_4)$ and $f(p_2)$ can be interchanged. By symmetry, the argument follows a similar pattern if $f(p_3)$ were chosen to be on the bottom or a side of the square S_2. One can see directly that the only mappings of a γ-star into a γ-star which preserve the given distances (1) are translations followed by one of the eight elements of the group of the square centered at $(x', y' + \gamma)$. We shall refer to this group as G.

γ-stars are "small" sets on which ρ-isometries must be Euclidean. They are important for two other reasons: first, any ρ-isometry on a γ-star can be uniquely extended to a certain unbounded set; second, any set of positive Lebesgue measure contains many γ-stars.

LEMMA 2. *Let S be the γ-star at (x, y) in $\mathbf{R}^2$ and $E \supseteq S$ any subset of $Q = \{(x, y) | |x - y| < \gamma\} \cup \{(x, y) | |x + y| < \gamma\}$. Then any ρ-isometry f on S has a unique extension to E. In particular, by Lemma 1, any ρ-isometry is Euclidean and belongs to G.*

Proof. First consider the geometry of the situation in FIG. 5, both for understanding and to clarify the notation. We split Q into five subsets:

$$\begin{aligned}
Q_1 &= \{(x, y) | |x - y| < \gamma\} \cap \{(x, y) | x + y > \gamma\} \\
Q_2 &= \{(x, y) | |x + y| < \gamma\} \cap \{(x, y) | y - x > \gamma\} \\
Q_3 &= \{(x, y) | |x - y| < \gamma\} \cap \{(x, y) | x + y < -\gamma\} \\
Q_4 &= \{(x, y) | |x + y| < \gamma\} \cap \{(x, y) | y - x < -\gamma\} \\
Q_5 &= \{(x, y) | |x| + |y| \leqslant \gamma\}.
\end{aligned} \tag{2}$$

Every point p in, for example, Q_1 is at the intersection of the vertical side of a square centered at q_2 and the horizontal side of a square centered at q_1 as illustrated in FIG. 5(ii). Only p' in Q_5 has that same set of ρ-distances from q_1 and q_2; p and p' have different ρ-distances from q_3 as well as from q_4. Points on the line segment between q_1 and q_2 are uniquely located by their ρ-distances from q_1 and q_2. Any point p in all of Q is at the intersection of a horizontal and a vertical side of squares centered at two adjacent points of q_1, q_2, q_3 and q_4. p can be

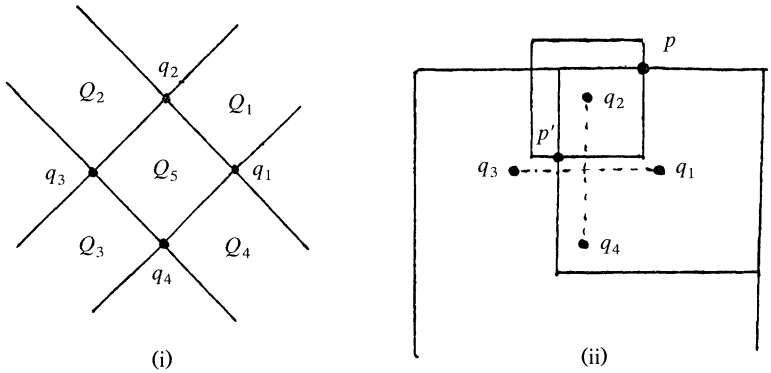


FIG. 5. (i) The set Q. (ii) p is uniquely located in terms of its distances from points of a γ-star.

located uniquely by specifying its ρ-distances from these four points, which are fixed throughout and determine Q.

For calculation, these ρ-distances can be given in terms of the coordinates of the point $p = (x, y)$. Associate with any point in $Q = Q((a, b), \gamma)$ a four-tuple $[\rho_1, \rho_2, \rho_3, \rho_4]$ of its ρ-distances from the defining points of Q: q_1, q_2, q_3, q_4.

If the ρ-isometry f is simply a translation, then the assertion is immediate in view of the discussion above. For ease of computation and without loss of generality, let $(a, b) = (0, 0)$; that is, consider f modulo a translation of E. $p \in E$ and the four-tuple $[\rho_1, \rho_2, \rho_3, \rho_4]$ in terms of the coordinates of p are

$$\begin{aligned}
&[y, x, \gamma + x, \gamma + y] \text{ if } p \in Q_1\\
&[\gamma - x, -x, y, \gamma + y] \text{ if } p \in Q_2\\
&[\gamma - x.\gamma - y, -y, -x] \text{ if } p \in Q_3\\
&[-y, \gamma - y, \gamma + x, x] \text{ if } p \in Q_4\\
&[\gamma - x, \gamma - y, \gamma + x, \gamma + y] \text{ if } p \in Q_5.
\end{aligned} \tag{3}$$

The eight possible ρ-isometries acting on the γ-star S are rotations 90°, 180°, 270°, and 360° counterclockwise; reflections across the lines $x = 0$, $y = 0$, $y = x$, and $y = -x$. $p \in Q_1$ has the set of ρ-distances $[y, x, \gamma + x, \gamma + y]$ from the points of S; $p_i = q_i$, $i = 1, 2, 3, 4$. If f is to be extended to an isometry on p consistent with f on S, then $f(p)$ will have the same ρ-distances from points of $f(S)$. If f is a 90° rotation counterclockwise then $f(p_1) = q_2$, $f(p_2) = q_3$, $f(p_3) = q_4$ and $f(p_4) = q_1$. We have $\rho_4 = \rho(p, q_4) = \rho(p, p_4) = \rho(f(p), f(p_4)) = \rho(f(p), q_1)$. This last is now the first entry in the four-tuple of ρ-distances of $f(p)$ from q_i, $i = 1, 2, 3, 4$, yielding $[\gamma + y, y, x, \gamma + x]$, which is the unique set of ρ-distances for a point in Q_2 with coordinates $(-y, x)$. This represents a 90° rotation counterclockwise of $p = (x, y)$.

Another case for $p \in Q_5$ and f on S is a reflection across $y = x$. The ρ-distances of p from points of S will be $[\gamma - x, \gamma - y, \gamma + x, \gamma + y]$. $f(p_1) = q_2$, $f(p_2) = q_1, f(p_3) = q_4, f(p_4) = q_3$, so that $f(p)$ has the four-tuple of ρ-distances $[\gamma - y, \gamma - x, \gamma + y, \gamma + x]$. $f(p)$ is uniquely located in Q_5 with coordinates (y, x).

The verification of Lemma 2 for the other cases is straightforward, following the same ideas as the examples chosen and using the symmetry properties of the square.

If $p \notin Q$, say in $\overline{Q}_a$, where $Q_a = \{(x, y) | x + y > \gamma \text{ and } x - y > \gamma\}$, its distances from points of S are given by $[x - \gamma, x, x, x + \gamma]$. Any point on the vertical line $x = k, |y| \leqslant k - \gamma$, has the same set of distances from S. Points on that line segment cannot be distinguished from one another in terms of their distances from points of S. A certain amount of vertical (or horizontal) movement cannot be detected from points within the convex hull of the γ-star.

A more concise proof of the analogy of Lemma 2 in $\mathbf{R}^n$ for $n \geqslant 2$ is possible if attention is restricted to the convex hull of the n-dimensional γ-star. See Lemma 2 of [3]. This is sufficient to answer the questions of Schattschneider [6] as to the nature of the group of isometries of $\mathbf{R}^n$ with respect to the max metric. Any two points $p, q \in \mathbf{R}^n$ belong to the convex hull of some γ-star S. By Lemma 2 cited above, any ρ-isometry of $\mathbf{R}^n \to \mathbf{R}^n$ must map the convex hull of S in the same way as it maps the points of S; namely a translation followed by a transformation of the group of the n-cube, which maps points of an n-dimensional γ-star to other such points, so that Euclidean distances are preserved. To relate the group of isometries of $(\mathbf{R}^n, d_T)$ to that of $(\mathbf{R}^n, \rho)$ we cite here a theorem from [2]:

THEOREM 1. *If U_1 and U_2 are the unit spheres of the n-dimensional Minkowski spaces M_1 and M_2, then any linear transformation which maps U_1 onto U_2 is an isometry of M_1 onto M_2.*

These results can be generalized for a Minkowski metric whose unit "circle" is given by a closed centrally symmetric polygon. For example, with a hexagon of side γ the analog of a γ-star would be the set of six corner points H. A point of H has distance γ to an adjacent point, distance 2γ to all other points of H. Every point inside the hexagon can be located uniquely in terms of its distances from points of H. Any ρ-isometry (ρ defined by a hexagon metric now) must carry points of H to points of H via some Euclidean transformation, a translation followed by a motion which is a symmetry of the hexagon.

4. ρ-isometries and extensions on more general sets. Returning to ρ-isometries with the max metric, if E is an open set in the usual topology, there is a γ-star $S \subseteq E$ centered at every point $p \in E$. Any ρ-isometry acting on E must be locally Euclidean. If E is open and connected, then f is the same Euclidean isometry on all of E. Any two points p and q in E can be connected by a simple arc, which can be covered by a chain of finitely many overlapping closed convex hulls of γ-stars contained in E. Starting at p and proceeding from set to set in the chain, the ρ-isometry at q must have the same action on q as on p.

Looking back at Example 3 and forward to Example 5, one can see that it is possible for a ρ-isometry to act like a different Euclidean isometry on different parts of a set. It might seem that a γ-star configuration is too specific to be useful in studying the action of a ρ-isometry on more general sets in $\mathbf{R}^2$ or $\mathbf{R}^n$. It turns out, however, that at almost every point x of a measurable set E there are infinitely many γ-stars of arbitrarily small size centered at x. Thus, when considering the action of a ρ-isometry f on E, we can look locally at its action on a γ-star. m denotes linear Lebesgue measure, and $\lambda = m \times m$.

LEMMA 3. [3] *E is a measurable set in $\mathbf{R}^2$, $\lambda(E) > 0$, and $\delta > 0$. At almost every point $p \in E$ there is a γ-star at p contained in E with $0 < \gamma < \delta$.*

The proof uses the Lebesgue density theorem and the Fubini Theorem. The remaining points (x, y) are those which are linear m-density points on both the horizontal E^y and vertical E_x sections at (x, y) in E. This result shows that there is a kind of local geometric symmetry on an arbitrarily small scale at almost every point in a measurable set.

The type of set we may consider can be widened beyond the class of measurable sets. By standard methods it is easy to show that a ρ-isometry f on a set E may be uniquely extended to a ρ-isometry $\bar{f}$ on the closure of E so that $\bar{f}/_E = f$. Then if $\lambda(\overline{E}) > 0$, the following theorem describes a ρ-isometry on E, except on a set of measure zero.

THEOREM 2. [3] *$E \subseteq \mathbf{R}^2$, $\lambda(\overline{E}) > 0$. There exist pairwise disjoint closed squares, sides of slope $\pm 1, \{S_i\}_{i=1}^{\infty}$, such that for any ρ-isometry f of E and any $i \geqslant 1$ the restriction of f to $E \cap S_i$ is Euclidean belonging to G. $\lambda(\overline{E} \setminus \bigcup_{i=1}^{\infty} S_i) = 0$.*

The proof, which holds in $\mathbf{R}^n$, employs the Vitali covering theorem. The closed squares are closed convex hulls of γ-stars in $\overline{E}$. On "most" of a set, a ρ-isometry must be locally Euclidean.

There are many examples of ρ-isometries where the set E has countably many disjoint parts and the ρ-isometry is a different Euclidean isometry on each.

EXAMPLE 5. In $\mathbf{R}^2$ consider the family of isosceles triangles $T_i, i = 0, 1, 2, \ldots$ with vertices $(3i, 0), (3i + 1, 1), (3i + 2, 0)$. Include the interior points. Let f be a transformation whose action is given by the sequence $(\delta_0, \delta_1, \delta_2, \ldots), \delta_i = 0$ or 1 where 1 in the ith place represents leaving T_i fixed and 0 represents reflection of T_i across the x-axis. Any such f is a ρ-isometry on $\bigcup_{i=0}^{\infty} T_i = T$. Also, given any sequence of real numbers $(t_0, t_1, t_2, \ldots)$ where $t_0 = 0$ and $|t_{i+1} - t_i| \leqslant 1$, a ρ-isometry g can be defined so that the quantity t_i represents a vertical translation of the triangle T_i. f and g may be composed, resulting in a combination of local reflections and translations.

Here E is an unbounded set, but the example can be revised so that the bases and heights of the triangles are terms of a convergent series, and the resulting set $\bigcup_{i=0}^{\infty} T_i$ will be bounded. This sort of example can be constructed for a connected set: for $i = 0, 1, 2, \ldots$ let T_i' be the triangle with vertices $(2i, 0), (2i + 1, 1)$, $(2i + 2, 0)$ and its interior points; as above, define the action of the ρ-isometry f by the sequence $(\delta_0, \delta_1, \delta_2, \ldots)$. Again, the construction can easily be revised so that $T' = \bigcup_{i=0}^{\infty} T_i'$ is bounded as well as connected. Furthermore, sets of measure zero such as line segments can be attached. When placed appropriately, the ρ-isometry may be locally non-Euclidean on these while remaining locally Euclidean on the "substantial" part.

Remark. Two sets E and F are said to be equivalent by enumerable decomposition, in the sense of Banach and Tarski, if each can be partitioned into disjoint subsets $E_1, E_2, \ldots$ and $F_1, F_2, \ldots$ so that E_i is e-congruent to F_i for all $i \geqslant 1$; [1]. The ρ-congruent sets $T, f(T), g(T)$ and $f \circ g(T)$ are equivalent by enumerable decomposition, as are T' and $f(T')$, but the ρ-congruent sets of Example 1 are not.

On what types of sets can a ρ-isometry be uniquely extended to a ρ-isometry $\mathbf{R}^2 \to \mathbf{R}^2$? No bounded set will suffice.

PROPOSITION 2. *$E \subset \mathbf{R}^2$ is bounded and f is the identity on E. Then there is a point $p \notin \overline{E}$ and a ρ-isometry $\hat{f}$, $\hat{f}|_E = f$ and $\hat{f}$ is not Euclidean on $E \cup \{p\}$.*

Proof. E and $\overline{E}$ are bounded, so there is a square with sides of slope ± 1 enclosing $\overline{E}$. Let $q = (x_0, y_0)$ be its highest point. With $k > 0$, choose $p = (x', y_0 + k)$ with $x_0 < x' < x_0 + k$. Define $\hat{f} = f$ on E and $\hat{f}(p) = \hat{f}(x', y_0 + k) = (\frac{1}{2}(x_0 + x' + k), y_0 + k)$. $\hat{f}$ on $\{p\}$ is a small horizontal displacement. $\hat{f}$ is a ρ-isometry and not an e-isometry on $E \cup \{p\}$. There exist non-Euclidean extensions to much larger sets, for example to $\overline{Q_u(q)}$ and $\overline{Q_d(q')}$ where q' is the vertex opposite to q. Define $\hat{f}$ as the identity on E and as a reflection about the line through q and q' for points in $\overline{Q_u(q)}$ and $\overline{Q_d(q')}$.

On the other hand, a sparse set E which is non-dense and of measure zero can have the property that a ρ-isometry defined on E is uniquely extendable to a ρ-isometry of $\mathbf{R}^n \to \mathbf{R}^n$. One example in $\mathbf{R}^2$ is the set $E = \{(0, n), (m, 0) | m, n \in \mathbf{Z}\}$ which contains a γ-star for arbitrarily large γ. This condition is sufficient but not necessary, e.g., a ρ-isometry on an infinite vertical strip of positive width must be Euclidean.

If the boundary of the convex set defining the Minkowski metric contains a straight line segment, it will always be the case that for a bounded set E a non-Euclidean extension of the identity can be constructed as above. All points on some line segment outside E and parallel to the line segment in the unit "circle" have the same distance from a point in E.

5. Other four-point sets on which ρ-isometries must be Euclidean. It is an interesting question whether γ-stars are the only four-point sets in $\mathbf{R}^2$ on which ρ-isometries must be Euclidean. The answer is no: a variation on a γ-star produces such a set, but the variation must be done symmetrically. One type of four point set on which ρ-isometries must be Euclidean will be called a skew γ-star at p (see FIG. 6). The four points are centered at $p = (x, y)$ with coordinates $p_1 = (x + \gamma, y - \varepsilon)$ $p_2 = (x + \varepsilon, y + \gamma)$ $p_4 = (x - \varepsilon, y - \gamma)$ $p_3 = (x - \gamma, y + \varepsilon), |\varepsilon| < \gamma$. (Call ε the eccentricity of the γ-star.) ρ-distances to adjacent points are all $\gamma + |\varepsilon|$, and 2γ to opposite points. Notice that the convex hull of a skew γ-star is a tilted square.

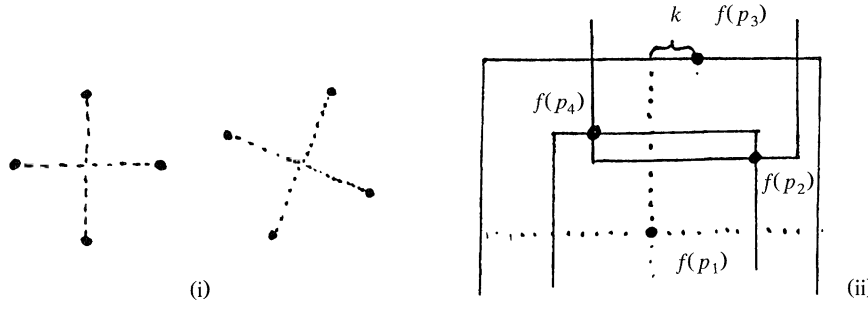


FIG. 6. (i) a γ-star and a skew γ-star. (ii) $f(p_2) = (k - \gamma - \epsilon, \gamma + \epsilon)$ and $f(p_4) = (\gamma + \epsilon, \gamma - \epsilon)$.

PROPOSITION 3. *A ρ-isometry f on a skew γ-star must be Euclidean.*

Proof. Without loss of generality, $\varepsilon > 0$. $f(p_1)$ is a translation of p_1 to $(0,0)$. $f(p_3)$ must be on a square S_1 of side 4γ centered at $f(p_1)$. Suppose it is on the top side. $f(p_2)$ and $f(p_4)$ are on a square S_2 of side $2\gamma + 2\varepsilon$ centered at $f(p_1)$ also; they are located as well on another square S_3 of the same size centered at $f(p_3)$. Where on the top side of S_1 must $f(p_3)$ be? One can eliminate directly the points at the corners or the center of the top of S_1. Let $f(p_3)$ be an e-distance $k > 0$ to the right of center (see FIG. 6).

Now $\rho(f(p_2), f(p_4)) = 2\gamma = \max(2\gamma + 2\varepsilon - k, 2\varepsilon)$. This implies $k = 2\varepsilon$ and the points $\{f(p_i)\}_{i=1}^4$ are in the configuration of skew γ-stars. The other cases are similar.

Modulo translations, there are sixteen ρ- (and e-) congruences for any skew γ-star. Each is an element from the group of the square or such an element composed with a reflection across one of the lines passing through opposite points of the skew γ-star. The ρ-isometry mapping the skew γ-star at ρ with eccentricity $\varepsilon > 0$ to the skew γ-star at ρ with eccentricity $-\varepsilon$ is equivalent to a (Euclidean) rotation through $2 \arctan \varepsilon/\gamma$. Therefore, for any angle of rotation, there is a skew γ-star for which that rotation is a ρ-isometry.

The other type of set on which a ρ-isometry must be Euclidean is the set of vertices of a rectangle with sides of slope ± 1. In a sense, the γ-star is the limiting case of all these sets. For brevity, denote the collection of all such sets, including γ-stars, as **S**.

The case of four points whose convex hull is a rectangle F, sides of slope ± 1, is treated in a way which points out an important feature of the max metric in $\mathbf{R}^2$. q_0 is said to be between q_1 and q_2 iff $\rho(q_1, q_2) = \rho(q_1, q_0) + \rho(q_0, q_2)$. The set of points satisfying this equation is denoted by $B(q_1, q_2)$. Geometrically, it is the rectangle and its interior with sides of slope ± 1 such that the line segment from q_1 to q_2 is the diagonal. If $q_1', q_2' \in B(q_1, q_2)$ then $B(q_1', q_2') \subseteq B(q_1, q_2)$ with equality only if q_1' and q_2' are coincident with q_1 and q_2 or are the other vertices of $B(q_1, q_2)$.

PROPOSITION 4. *A ρ-isometry f on the vertices of a rectangle F whose sides have slope ± 1 must be Euclidean.*

Proof. If $\{p_1, p_2, p_3, p_4\} = P$ are vertices of the rectangle F, counterclockwise, then $F = B(p_1, p_3) = B(p_2, p_4)$. We have

$$\begin{aligned} &\rho(p_1, p_2) = \rho(p_3, p_4) = a; \qquad \rho(p_2, p_3) = \rho(p_1, p_4) = b \\ &\rho(p_2, p_4) = \rho(p_1, p_3) = a + b \qquad \text{where } a, b > 0. \end{aligned} \tag{4}$$

f is a ρ-isometry on P. $f(p_2)$ and $f(p_4)$ belong to $B(f(p_1), f(p_3))$, so the latter must contain $B(f(p_2), f(p_4))$. $f(p_1)$ and $f(p_3)$ belong to $B(f(p_2), f(p_4))$ as well, so the two sets are equal and $f(p_i)$, $i = 1, 2, 3, 4$ are vertices of a rectangle, with sides of slope ± 1, of the same dimensions (4) as F, and f will be Euclidean.

6. The separating set of Schoenberg. To establish that the class **S** contains the only four-point sets and thus the minimal sets on which ρ-isometries must be Euclidean, we use a construction from [7] attributed by the author to I. J. Schoenberg. This is perhaps a specialized result, but its proof illustrates the

usefulness of the Schoenberg construction for the study of ρ-isometries on finite sets in $\mathbf{R}^2$. One can construct separating sets in $\mathbf{R}^2$ for some other Minkowski metrics as well.

Definition of a separating set: E is a finite set with an even number of points $2m$. Consider first the lines of equation $y = x + s$ such that m of the points lie above the line and m below. $y = x + s_1$ and $y = x + s_2$ are the upper and lower bounds of such lines. Likewise, for lines $y = -x + t$, find the lower bound $y = -x + t_1$ and the upper bound $y = -x + t_2$. The closed rectangle formed by these four lines and its interior is called the separating set of E. See FIG. 7. Notice several useful facts: For p_0 in the separating set

$$\rho(p_a, p_c) = \rho(p_a, p_0) + \rho(p_0, p_c), \quad \text{for } p_0 \in B(p_a, p_c), \qquad p_i \in Q_i. \quad (5)$$

ρ-distances between points in Q_a and Q_c and between points in Q_b and Q_d are their horizontal and vertical e-distances respectively. Many different sets will have the same separating set. There are no points of E in the two strips whose intersection is the rectangle; otherwise the lines would be positioned differently. For the same reason when E consists of four points, it will not be the case that adjacent open quadrants, e.g., Q_a and Q_d, both contain points of E.

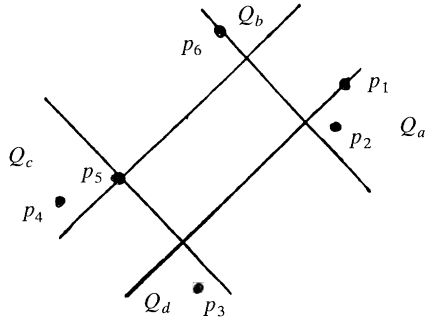


FIG. 7. The separating set of $P = \{p_1, p_2, p_3, p_4, p_5, p_6\}$.

PROPOSITION 5. *For any four-point set* $E = \{p_1, p_2, p_3, p_4\}$ *which is not of the class* **S**, *there is a* ρ*-isometry on* E *which is not Euclidean.*

Proof. We construct a non-Euclidean ρ-isometry for any $E \notin \mathbf{S}$. Treat the simpler cases first. (a) If all four points are on a line of slope $+1$ or -1, use the inverse of the ρ-isometry of Example 2. (b) If three points are on a line L of slope $+1$ or -1, construct line L' through p_4 perpendicular to L. $q = L \cap L'$. Half-pivot about q, bringing p_4 onto L and changing its e-distances to at least one of p_1, p_2, and p_3. In what follows, L_i and L_i' are lines of slope ± 1 containing $p_i \ni L_i \perp L_i'$ for $i = 1, 2, 3, 4$. (c) Two points p_1 and p_2 are on a line L of slope $+1$ or -1.

(i) The other two points are to one side of L.

If there is a line $L' \perp L$ so that three of the points lie to one side of L' and $L' \cap L = q$ is strictly between p_1 and p_2, then half-pivot about q, bringing one of p_1 or p_2 onto L' and leaving the other on L, changing e-distance.

If there is no such line as L' above, see FIG. 8, p_3 lies on or above $L_2', L_2' \perp L, p_2 = L_2' \cap L$ and p_4 lies below $L_1', L_1' \perp L, p_1 = L_1' \cap L$. L_3 and L_4 are parallel to L, with $p_3 \in L_3$ and $p_4 \in L_4$. One of those lines lies between L and the other; let it be L_3. $q = L_1' \cap L_4$ has three points of E in one quadrant and the fourth in the opposite. A half-pivot about q puts p_1 and p_4 on the same line L_1', changing e-distance. For other configurations, one can always choose one of $L_1' \cap L_3, L_1' \cap L_4, L_2' \cap L_3$ or $L_2' \cap L_4$ as a half-pivot point.

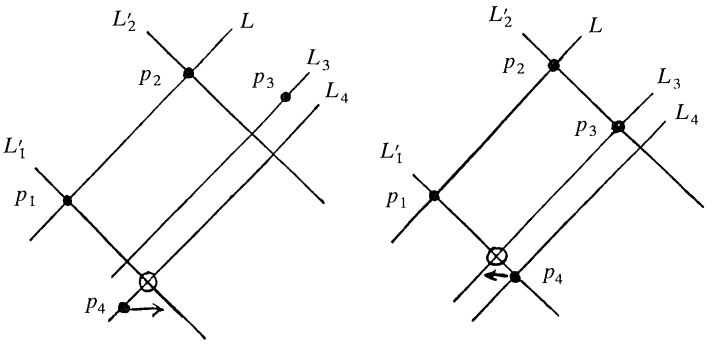


FIG. 8. Case c(i) of Proposition 5.

If both pairs p_1, p_2 and p_3, p_4 are on parallel lines, of slope $+1$ or -1, and if the projection of the segment from p_1 to p_2 onto L_3 would be properly contained in the segment from p_3 to p_4, then use p_3 and p_4 as the beginning pair and there will be a line L' as above.

(ii) Two points p_1 and p_2 are on a line L of slope $+1$ or -1; p_3 is on one side of L, p_4 is on the other. $L = L_1 = L_2$.

If there is a line $L' \perp L$ so that p_3 is on one side of L' and p_4 is on the other and if $q = L' \cap L$ is strictly between p_1 and p_2, there are two points of E in one quadrant of q and two in the opposite. Half-pivot about q, bringing p_2 onto L'.

If L_3' and L_4' lie outside and both above or both below the segment from p_1 to p_2 then construct lines L_3 and L_4 through p_3 and p_4 parallel to L. Suppose L_3' lies between L_4' and L_2'. $q = L \cap L_4'$. One closed quadrant of q contains three points of E, and p_4 is in the opposite closed quadrant. Half-pivot about q and bring p_4 onto L. If $L_3' = L_4', L_2' = L_3'$, or $L_3' = L_4' = L_2'$, let $q = L \cap L_4'$ also.

(d) Only one point of E is on any one line of slope $+1$ or -1. Construct the separating set of E; there are four cases where a half-pivot point can be chosen easily, as in FIG. 9. Neither these points nor those of cases (b) and (c) are unique.

Consider now a four-point set E, not of types already studied, where there is only one point of E on each line of the separating set S_0 and none at any vertex. See FIG. 10. Without loss of generality, the center of S_0 is at $(0, 0)$. The four lines bounding S_0 are $y = x \pm b$ and $y = -x \pm a$, taking $a > b > 0$. The coordinates

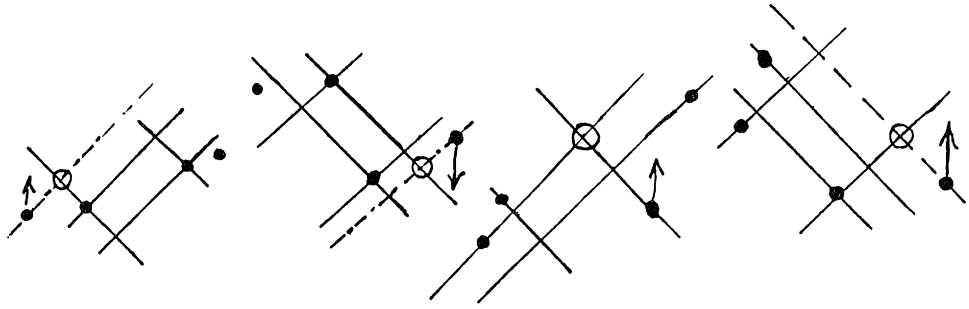


FIG. 9. Left/right half-pivot points are circled. Other possible configurations are obvious variations of these.

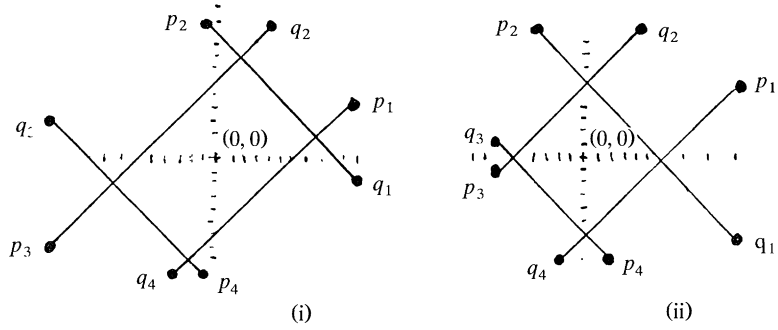


FIG. 10. (i) $a > b > 0$. (ii) $a = b$.

of the vertices of S_0 are

$$v_1 = \left(\frac{a+b}{2}, \frac{a-b}{2}\right) \qquad v_3 = \left(\frac{-(a+b)}{2}, \frac{-(a-b)}{2}\right)$$
$$v_2 = \left(\frac{a-b}{2}, \frac{a+b}{2}\right) \qquad v_4 = \left(\frac{-(a-b)}{2}, \frac{-(a+b)}{2}\right).$$

$\{z_i\}_{i=1}^4$ are positive numbers; then the set E must have the configuration of the points $\{p_i\}_{i=1}^4$ or $\{q_i\}_{i=1}^4$ with $e(p_i, v_i) = e(q_i, v_i) = \sqrt{2}z_i$, $i = 1, 2, 3, 4$ where the coordinates are given as in FIG. 10.

$$\begin{aligned}
p_1 &= \left(\frac{a+b}{2} + z_1, \frac{a-b}{2} + z_1\right) & q_1 &= \left(\frac{a+b}{2} + z_1, \frac{a-b}{2} - z_1\right)\\
p_2 &= \left(\frac{a-b}{2} - z_2, \frac{a+b}{2} + z_2\right) & q_2 &= \left(\frac{a-b}{2} + z_2, \frac{a+b}{2} + z_2\right)\\
p_3 &= \left(\frac{-(a+b)}{2} - z_3, \frac{-(a-b)}{2} - z_3\right) & q_3 &= \left(\frac{-(a+b)}{2} - z_3, \frac{-(a-b)}{2} + z_3\right)\\
p_4 &= \left(\frac{-(a-b)}{2} + z_4, \frac{-(a+b)}{2} - z_4\right) & q_4 &= \left(\frac{-(a-b)}{2} - z_4, \frac{-(a+b)}{2} - z_4\right)
\end{aligned} \tag{6}$$

Keep in mind, for example, that $Q_r(p_1)$ and $Q_r(p_4)$ are both contained in $Q_r(p_3)$; then $\rho(p_1, p_3)$ and $\rho(p_4, p_3)$ are the differences in the x-coordinates of the

points, and the x-coordinates of p_1 and p_4 are both larger than that of p_3. $\rho(p_i, p_j) = \rho(q_i, q_j)$ for $i, j = 1, 2, 3, 4$. Namely,

$$\begin{aligned} \rho(p_1, p_3) &= a + b + z_1 + z_3 & \rho(p_1, p_4) &= a + z_1 + z_4 \\ \rho(p_2, p_4) &= a + b + z_2 + z_4 & \rho(p_2, p_3) &= a + z_2 + z_3 \\ \rho(p_1, p_2) &= b + z_1 + z_2 & \rho(p_3, p_4) &= b + z_3 + z_4. \end{aligned} \tag{7}$$

Therefore, $f(p_i) = q_i$, $i = 1, 2, 3, 4$ is a ρ-isometry on E. f is not Euclidean, for $e(p_1, p_3) \neq e(q_1, q_3)$ unless $a = b$. $e(p_1, p_2) \neq e(q_1, q_2)$ unless $z_1 = z_2$. This can be checked by the ordinary distance formula in $(\mathbf{R}^2, e)$. One can also see this geometrically by considering the legs of the two right triangles for which each of those e-lengths is the hypotenuse. In general, f will preserve e-distances iff the separating set is a square and $z_1 = z_2 = z_3 = z_4$. This is another characterization of a skew γ-star. If E has the configuration $\{p_i\}_{i=1}^4$, apply ρ-isometry f; if E has the configuration $\{q_i\}_{i=1}^4$, apply ρ-isometry f^{-1}. The case $b > a > 0$ is analogous. f is a variation of the ρ-isometry employed in Lemma 1 of [7].

Remark. The action of f could be thought of as simultaneous half-pivots about the points v_1, v_2, v_3 and v_4. For any point within the rectangle the ρ-distances to the points $\{p_i\}_{i=1}^4$ are the same as those to $\{q_i\}_{i=1}^4$. In other words, the ρ-isometry on the set E can be extended to the rectangle S_0 by defining f there as the identity map. f can also be extended to the rays of slope ± 1 from each vertex in which p_i lies, $i = 1, 2, 3, 4$. Each ray is mapped to the ray of opposite slope which contains v_i and q_i, $i = 1, 2, 3, 4$.

7. Some further questions. The non-Euclidean ρ-isometries of Examples 1 and 4 and the extended f of the above Remark have a maximal set beyond which they cannot be extended. Does each such ρ-isometry have a maximal set and can these maximal sets be characterized? Interesting combinatorial problems are encountered in looking for minimal sets on which ρ-isometries must be Euclidean in $(\mathbf{R}^n, \rho)$, $n > 2$, or for other Minkowski metrics where the unit "circle" is a polygon in $\mathbf{R}^2$. Lastly, it would be instructive to explore specific examples of set isometries for Minkowski metrics where the unit "circle" contains arcs of ordinary circles as well as straight line segments.

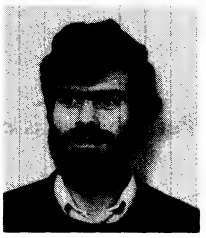
REFERENCES

1. S. Banach and A. Tarksi, Sur la décomposition des ensembles de points en parties respectivement congruentes, *Fundamenta Mathematicae*, 6 (1924) 244–277.
2. Russell V. Benson, Euclidean Geometry and Convexity, McGraw-Hill, 1966, p. 241.
3. Regina Cohen and James W. Fickett, The max norm in $\mathbf{R}^n$, isometries and measure, *Colloquium Mathematicum*, 47 (1982) 17–22.
4. Richard S. Millman and George D. Parker, Geometry, a Metric Approach with Models, Springer-Verlag, New York, 1981.
5. Nathan Jacobson, Basic Algebra I, W. B. Freeman, San Francisco, 1974, p. 351.
6. Doris J. Schattschneider, The taxicab group, this MONTHLY, 91 (1984) 423–428.
7. Dorothy Wolfe, Imbedding a finite metric set in an n-dimensional Minkowski space, Koninklijke Nederlandse Akademie van Wetenschappen, *Proceedings*, *Series A*, LXX (1967) 136–140.

On Stability of Invariant Subspaces of Matrices

A. C. M. RAN, *Vrije Universiteit, Amsterdam, The Netherlands*

L. RODMAN, *The College of William and Mary*



ANDRÉ C. M. RAN received his M.Sc. and his Ph.D. at the Vrije Universiteit in Amsterdam. In both cases his thesis advisors were M. A. Kaashoek and I. Gohberg. After receiving his Ph.D. in 1984 he spent a year in the US, mainly at Virginia Tech, on a grant from the Niels Stensen Stichting (Amsterdam). Since 1985 he has been working at the Vrije Universiteit again. His research interests are operator theory, matrices, and control theory.



LEIBA RODMAN received his Ph.D. in mathematics from Tel-Aviv University in 1978. Since then he held permanent positions at Tel-Aviv University (Israel), Arizona State University and The College of William and Mary (Virginia), and visiting positions at University of Calgary (Canada), Vrije Universiteit (Amsterdam) and University of California (San Diego). His research interests are mainly in linear mathematics (operators, matrices, linear systems and control).

1. Introduction. In this paper we study the behavior of invariant subspaces of matrices, under small additive perturbations of these matrices. In particular, we will be interested in studying the rate of convergence (in terms of the norm of the perturbation) of the invariant subspace of the perturbed matrix. It turns out that this rate of convergence can be different from that of eigenvectors only. In some cases the rate of convergence of invariant subspaces is actually much better than that of eigenvectors if one has the choice of an invariant subspace for the perturbed matrix. Also, new phenomena appear; for example, the rate of convergence can be different if all nearby invariant subspaces of the perturbed matrix are considered. In this paper we develop the concepts of stability that reflect the rate of convergence of invariant subspaces. We point out some new phenomena that occur in this situation and prove partial results concerning characterization of invariant subspaces with a given rate of convergence. Many questions here are open and we indicate some of them.

It is well known that the eigenvalues of an $n \times n$ (complex or real) matrix depend continuously on the matrix. The behavior of eigenvectors has been extensively studied from the point of view of applications in numerical analysis [**W, GvL**]. In fact, the behavior of eigenvectors can be basically described as follows: it falls in either one of the following three classes:

1) hopeless: consider, for example, the eigenvector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ of the matrix $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Perturbing the matrix to $\begin{bmatrix} 0 & 0 \\ \varepsilon & 0 \end{bmatrix}$ with ε a small number we see that the eigenvector

of the perturbed matrix is far away from $\begin{bmatrix}1\\0\end{bmatrix}$. Of course, there is nothing special about $\begin{bmatrix}1\\0\end{bmatrix}$; any other eigenvector of $\begin{bmatrix}0&0\\0&0\end{bmatrix}$ is just as hopeless.

2) continuous but bad: the standard example here is the $n \times n$ matrix

$$J_n = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}.$$

Perturbing this matrix with an $\varepsilon > 0$ in the left lower corner

$$J_n(\varepsilon) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ \varepsilon & 0 & \cdots & \cdots & 0 \end{bmatrix},$$

we have the following eigenvectors of $J_n(\varepsilon)$: $y_j(\varepsilon) = (1, \varepsilon_j, \varepsilon_j^2, \ldots, \varepsilon_j^{n-1})^T$, $j = 1, \ldots, n$ where ε_j is given by $\varepsilon_j = |\varepsilon|^{1/n}\exp(2\pi ij/n)$. Clearly, for the vectors y_j we have $\|y_j - e_1\| = 0(|\varepsilon|^{1/n})$ where $e_1 = (1, 0, \ldots, 0)^T$ is the eigenvector of J_n. (Here and everywhere in the paper we use the Euclidean norm for the column vectors.) We call it bad because the rate of convergence of the eigenvectors of $J_n(\varepsilon)$ to e_1, is $0(|\varepsilon|^{1/n})$, which is slow for large n.

3) continuous and good: the case of an $n \times n$ matrix with n different and well-separated eigenvalues. In this case, as we shall also see later, the perturbation of the eigenvectors is of the same order of magnitude as the perturbation of the matrix.

It is natural to classify along these lines the behavior of general invariant subspaces. (Clearly, an eigenvector can be identified with the one-dimensional subspace it generates). Given an $n \times n$ complex matrix A, a subspace $M \subset \mathbb{C}^n$ is called *A-invariant* if $AM \subset M$. Here $\subset$ stands for (not necessarily proper) inclusion. Every invariant subspace is spanned by a set of eigenvectors and generalized eigenvectors of A (recall that $\chi \neq 0$ is a generalized eigenvector of A belonging to the eigenvalue of λ if $(A - \lambda I)^j\chi = 0$ for some $j > 1$).

In the next section we shall recall some of the results on stability of invariant subspaces, before going to the main topic of this paper, which is treated in sections three and four.

2. Stable invariant subspaces. The stability properties of invariant subspaces under perturbations of the matrix have been studied extensively in the last decade **[BGK, CD, KvdMR, GLR2, GR]**. A source where all this material is gathered (and which includes much more) is **[GLR2]**. We start with the definition of stable invariant subspaces. We shall denote the set of invariant subspaces of a matrix A by $\mathrm{Inv}(A)$.

Let A be an $n \times n$ complex matrix and let $M \in \mathrm{Inv}(A)$. M is called *stable* if for all $\varepsilon > 0$ there is a $\delta > 0$ such that $\|B - A\| < \delta$ implies the existence of $N \in \mathrm{Inv}(B)$ with $\|P_M - P_N\| < \varepsilon$. Here P_M (resp., P_N) is the orthogonal projection

on M (resp., N), and $\|X\|$ denotes the operator norm of a matrix X, i.e., $\|X\| = \max_{\|\chi\|=1}\|X\chi\|$, with $\|\chi\|$ being the usual Euclidean norm, $\|(\chi_1, \ldots, \chi_n)^T\|^2 = \sum_{j=1}^n |\chi_j|^2$. The main result on stability is the following theorem.

THEOREM 2.1. **[BGK, CD]**. *Let A be an $n \times n$ complex matrix and $M \in \mathrm{Inv}(A)$. Then M is stable if and only if for each eigenvalue λ of A with $\dim \mathrm{Ker}(A - \lambda I) > 1$ either $M \cap \mathrm{Ker}(A - \lambda I)^n$ is (0) or it is $\mathrm{Ker}(A - \lambda I)^n$. For eigenvalues λ of A with $\dim \mathrm{Ker}(A - \lambda I) = 1$ the space $M \cap \mathrm{Ker}(A - \lambda I)^n$ is an arbitrary A-invariant subspace contained in $\mathrm{Ker}(A - \lambda I)^n$.*

We denote by $\mathrm{Ker}\, X$ the kernel of the matrix X. Observe that $\mathrm{Ker}(A - \lambda I)^n$ is just the root subspace corresponding to λ, i.e., the subspace spanned by all eigenvectors and generalized eigenvectors belonging to the eigenvalue λ. Equivalently, $\mathrm{Ker}(A - \lambda I)^n$ is the largest A-invariant subspace N with the property that the restriction A/N has λ as its only eigenvalue.

This stability result just tells us when M is continuously dependent on A. In case M is stable both the "continuous and good" and the "continuous but bad" cases mentioned in the introduction can occur. To distinguish between the two we need another definition. An invariant subspace $M \in \mathrm{Inv}(A)$ is called *Lipschitz stable* if there exist positive numbers K and δ such that $\|B - A\| < \delta$ implies the existence of $N \in \mathrm{Inv}(B)$ with $\|P_M - P_N\| \leqslant K \cdot \|B - A\|$. Obviously, this defines precisely what we mean by "continuous and good." The following theorem describes such subspaces.

THEOREM 2.2 **[KvdMR]**. *Let A be an $n \times n$ complex matrix, let $M \in \mathrm{Inv}(A)$. Then M is Lipschitz stable if and only if for each eigenvalue λ of A either $M \cap \mathrm{Ker}(\lambda I - A)^n = (0)$ or $M \supset \mathrm{Ker}(\lambda I - A)^n$.*

As the proof is illustrative for our approach to another problem (in the next section) we shall give an outline of the proof here. In addition to **[GLR2]**, **[GK]** is an excellent source for the background theory used.

Proof. First suppose that M satisfies the condition of the theorem. Then M is a spectral subspace; i.e., there is a contour Γ in $\mathbb{C}$ such that $A - \lambda I$ is invertible for $\lambda \in \Gamma$ and

$$M = \mathrm{Im}\left[\frac{1}{2\pi i}\int_\Gamma (\lambda I - A)^{-1}\, d\lambda\right],$$

where we denote by $\mathrm{Im}\, X$ the range (i.e., the column space) of the matrix X. The matrix $(1/2\pi i)\int_\Gamma (\lambda I - A)^{-1}\, d\lambda$ is a (not necessarily orthogonal) projection. Now take δ small enough such that $\lambda I - B$ is invertible on Γ for $\|B - A\| < \delta$. Then put

$$N = \mathrm{Im}\left[\frac{1}{2\pi i}\int_\Gamma (\lambda I - B)^{-1}\, d\lambda\right].$$

The matrix $(1/2\pi i)\int_\Gamma (\lambda I - B)^{-1}\, d\lambda$ is again a projection. At this point we need the fact (the proof of which can be found, e.g., in **[GLR]**) that given subspaces $M, N \subset \mathbb{C}^n$ with orthogonal projection P_M and P_N on M and N, respectively, we have

$$\|P_M - P_N\| \leqslant \|Q_M - Q_N\|, \tag{2.1}$$

where Q_M and Q_N are arbitrary (not necessarily orthogonal) projections on M and

N. Using this fact we obtain

$$\begin{aligned}\|P_M - P_N\| &\leqslant \left\| \frac{1}{2\pi i} \int_\Gamma \left((\lambda I - B)^{-1} - (\lambda I - A)^{-1} \right) d\lambda \right\| \\ &= \left\| \frac{1}{2\pi i} \int_\Gamma (\lambda I - B)^{-1}(B - A)(\lambda I - A)^{-1}\, d\lambda \right\| \\ &\leqslant \frac{1}{2\pi} (\text{length of } \Gamma) \max_{\lambda \in \Gamma} \|(\lambda I - B)^{-1}\| \max_{\lambda \in \Gamma} \|(\lambda I - A)^{-1}\| \cdot \|B - A\|.\end{aligned}$$

So

$$\|P_M - P_N\| \leqslant K \|B - A\|$$

for all B sufficiently close to A where

$$K = \frac{1}{\pi} (\text{length of } \Gamma) \left(\max_{\lambda \in \Gamma} \|(\lambda I - A)^{-1}\|^2 \right).$$

So M is Lipschitz stable.

Conversely, suppose M is Lipschitz stable, but the condition in the theorem does not hold. It takes a little argument, which is left out here, to see that in that case also $M \cap \operatorname{Ker}(A - \lambda I)^n$ is Lipschitz stable for each eigenvalue λ. Hence, we can restrict our attention to the case $A = J_n$ is the $n \times n$ Jordan block with zero eigenvalue. Consider again the perturbation $J_n(\varepsilon)$ of J_n given by

$$J_n(\varepsilon) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ \varepsilon & 0 & \cdots & \cdots & 0 \end{bmatrix}.$$

Put $y_j = (1, \varepsilon_j, \ldots, \epsilon_j^{n-1})^T$, where ε_j is an nth root of $\varepsilon (j = 1, \ldots, n)$. For some $k < n$ we have $M = \operatorname{span}\{e_1, \ldots, e_k\}$ where e_j is the jth unit vector. Using the easily proved fact that M is a Lipschitz stable A-invariant subspace if and only if $M^\perp$ is a Lipschitz stable A^*-invariant subspace, we can assume without loss of generality that $k \leqslant n/2$. Take any k-dimensional invariant subspace N of $J_n(\varepsilon)$. It is spanned by k of the vectors y_j, and by renumbering the ε_j's we may assume $N = \operatorname{span}\{y_1, \ldots, y_k\}$. The subspace N is given by

$$N = \operatorname{Im} \begin{bmatrix} I_{k \times k} \\ \tilde{Y}_{n-m} Y_k^{-1} \end{bmatrix} = \operatorname{Im} \begin{bmatrix} Y_k \\ \tilde{Y}_{n-k} \end{bmatrix}$$

where

$$\begin{bmatrix} Y_k \\ \tilde{Y}_{n-k} \end{bmatrix} = [y_1 \ldots y_k]$$

is the $n \times k$ matrix with the vector y_j in its jth column. Thus Y_k is the $k \times k$

Vandermonde matrix with $(\varepsilon_1, \ldots, \varepsilon_k)$ as its second row:

$$Y_k = \begin{bmatrix} 1 & 1 & & 1 \\ \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_k \\ \vdots & \vdots & & \vdots \\ \varepsilon_1^{k-1} & \varepsilon_2^{k-1} & \cdots & \varepsilon_k^{k-1} \end{bmatrix}$$

At this point we need to compute the last column of Y_k^{-1}. It is easily done by using the formulas $Y_k^{-1} = (\det Y_k)^{-1}(\operatorname{adj} Y_k)$ and $\det Y_k = \prod_{i<j}(\varepsilon_j - \varepsilon_i)$. Further, computing the last column of $\tilde{Y}_{n-k}Y_k^{-1}$ one obtains that it equals $(p_1(\varepsilon_1, \ldots, \varepsilon_k), \ldots, p_{n-k}(\varepsilon_1, \ldots, \varepsilon_k))^T$ where $p_j(\chi_1, \ldots, \chi_k) = \Sigma \chi_1^{\alpha_1}\chi_2^{\alpha_2} \cdots \chi_k^{\alpha_k}$, where the sum is taken over all ordered k-tuples of non-negative integers $\alpha_1, \ldots, \alpha_k$, such that their sum is j. Put $u = (0, \ldots, 0, 1, p_1(\varepsilon_1, \ldots, \varepsilon_k), \ldots, p_{n-k}(\varepsilon_1, \ldots, \varepsilon_k))^T$ Then $u \in N$. Therefore,

$$\|P_M - P_N\|^2 \geqslant \|(P_M - P_N)u\|^2 \cdot \frac{1}{\|u\|^2} = \frac{1}{\|u\|^2}\|u - e_k\|^2$$

$$= \frac{1}{\|u\|^2} \cdot \sum_{j=1}^{n-k} |p_j(\varepsilon_1, \ldots, \varepsilon_k)|^2 \geqslant \frac{1}{4} \cdot \sum_{j=1}^{n-k} |p_j(\varepsilon_1, \ldots, \varepsilon_k)|^2 \quad (2.2)$$

for ε small enough. Suppose $p_j(\varepsilon_1, \ldots, \varepsilon_k) = 0$ for $j = 1, \ldots, n-k$. As $k \leqslant n/2$ we can form $\hat{Y}$, the matrix given by the first k rows of $\tilde{Y}_{n-k}$. Then, $\hat{Y}Y_k^{-1}$ has a zero last column, so $\det \hat{Y}Y_k^{-1} = 0$. But $\det \hat{Y} = \varepsilon_1^k \ldots \varepsilon_k^k \cdot \det Y_k$. So $\det \hat{Y}Y_k^{-1} \neq 0$. Hence, at least one of $p_1(\varepsilon_1, \ldots, \varepsilon_k), \ldots, p_k(\varepsilon_1, \ldots, \varepsilon_k)$ is non-zero. As $|p_j(\varepsilon_1, \ldots, \varepsilon_k)| = c|\varepsilon|^{j/n}$ in case $p_j(\varepsilon_1, \ldots, \varepsilon_k) \neq 0$ for some constant $c > 0$ independent of ε, this shows, in view of (2.2), that $\|P_M - P_N\|^2 \geqslant C_1|\varepsilon|^{k/n}$, where $C_1 > 0$ is independent of ε. It follows that M cannot be Lipschitz stable.

There is a remarkable property shared by stable and Lipschitz stable subspaces. Let A be an $n \times n$ complex matrix and $M \in \operatorname{Inv}(A)$. Let $\lambda_1, \ldots, \lambda_r$ be the different eigenvalues of A, and for each of these let Γ_j be a small contour around λ_j with λ_i outside Γ_j for $i \neq j$. Denote by P_j the spectral projection of A corresponding to Γ_j, i.e.,

$$P_j = \frac{1}{2\pi i}\int_{r_j} (\lambda I - A)^{-1}\, d\lambda.$$

The subspace M is called *strongly stable* if for each sequence $A_n \to A$ and *any* $M_n \in \operatorname{Inv}(A_n)$ with $\dim P_j M = \dim P_{nj}M_n$, (where P_{nj} is the spectral projection of A_n corresponding to Γ_j) we have

$$\|P_M - P_{M_n}\| \to 0, \quad n \to \infty. \quad (2.3)$$

Replacing (2.3) by

$$\|P_M - P_{M_n}\| \leqslant K\|A - A_n\|$$

for some $K > 0$ independent of n, we obtain the definition of *strong Lipschitz stability*. Roughly speaking, strong stability means that any sequence of subspaces, which has a chance of converging to M (in the sense of $\dim P_j M = \dim P_{nj}M_n$), really does converge to M.

As an $n \times n$ matrix has an invariant subspace of any dimension between 0 and n, clearly every strongly stable (or strongly Lipschitz stable) subspace is stable (or

strongly stable). Remarkably, the converse is also true. Inspection of the proof of Theorem 2.1 given in [**BGK**, **GLR2**] and of the proof of Theorem 2.2 yields

THEOREM 2.3. (a) $M \in \mathrm{Inv}(A)$ *is stable if and only if it is strongly stable.*
(b) $M \in \mathrm{Inv}(A)$ *is Lipschitz stable if and only if it is strongly Lipschitz stable.*

3. α-stability. In the previous section it became clear that Lipschitz stability is the key notion for understanding "continuous and good"-behaved stable invariant subspaces. However, there remains a gap between stable subspaces and Lipschitz stable subspaces. To better understand the behavior of stable invariant subspaces we need a more refined classification of stable invariant subspaces. Given $\alpha > 0$ and $M \in \mathrm{Inv}(A)$ a stable invariant subspace, M is called *α-stable* if there exist constants $K > 0, \delta > 0$ such that $\|B - A\| < \delta$ implies the existence of $N \in \mathrm{Inv}(B)$ with $\|P_M - P_N\| \leqslant K\|A - B\|^{1/\alpha}$. (Note that since $\|P_M - P_N\| \leqslant 2$ for any pair of subspaces $M, N \subset \mathbb{C}^n$ we can actually drop the requirement on B and A being sufficiently close, with the expense of replacing K by a bigger constant.) Clearly, if M is α-stable it is also β-stable with $\beta > \alpha$.

As in the cases discussed before we also introduce *strong α-stability*. Again, let $\lambda_1, \ldots, \lambda_\Gamma$ be the different eigenvalues of A, Γ_j contours around λ_j, with λ_i outside $\Gamma_j (i \neq j)$. Then a stable $M \in \mathrm{Inv}(A)$ is called *strongly α-stable* if for any sequence $A_n \to A$ and *any* $M_n \in \mathrm{Inv}(A_n)$ with $\dim P_j M = \dim P_{nj} M_n$ we have

$$\|P_M - P_{M_n}\| \leqslant K\|A - A_n\|^{1/\alpha}. \tag{3.1}$$

Roughly speaking, this means that any sequence M_n having a chance to converge to M does so and with rate specified by (3.1). Clearly, strong α-stability implies α-stability. Strong α-stability was studied in [**RR**] where the following result was proved:

THEOREM 3.1 [**RR**]. *Let A be an $n \times n$ matrix. Then $M \in \mathrm{Inv}(A)$ is strongly α-stable if and only if the following holds. The subspace $M \cap \mathrm{Ker}(A - \lambda I)^n$ is either* (0) *or* $\mathrm{Ker}(A - \lambda I)^n$ *for each eigenvalue λ of A satisfying one of the following two conditions*:
(a) $\dim \mathrm{Ker}(A - \lambda I) > 1$
(b) $\dim \mathrm{Ker}(A - \lambda I) = 1$ *and* $\alpha < \dim \mathrm{Ker}(A - \lambda I)^n$,
and for any eigenvalue λ not satisfying (a) or (b) $M \cap \mathrm{Ker}(A - \lambda I)^n$ is an arbitrary A-invariant subspace of $\mathrm{Ker}(A - \lambda I)^n$.

Several questions arise naturally in connection with the notions of α-stability and strong α-stability. First, are they equivalent? One might expect (in view of Theorem 2.3) that the answer is affirmative. However, we will produce an example showing that, in general, these notions are not equivalent. The next question then is: when does α-stability imply strong α-stability? We do not know the complete answer, but some observations concerning this question will be made later. Another problem is to find a description of all α-stable subspaces.

In case of eigenvectors (i.e., one-dimensional invariant subspaces) α-stability and strong α-stability are the same.

PROPOSITION 3.2. *Let A be an $n \times n$ matrix and let $M \in \mathrm{Inv}(A)$ be stable, and* $\dim M = 1$. *Then M is strongly α-stable if and only if M is α-stable.*

Proof. Let M be spanned by an eigenvector χ corresponding to an eigenvalue λ of A. If $\dim \operatorname{Ker}(\lambda I - A) > 1$ M cannot be stable, so necessarily $\operatorname{Ker}(\lambda I - A) = M$. By a standard reduction we may assume $\lambda = 0$, and $A = J_n$ is the $n \times n$ Jordan block with zero eigenvalue. Take $A(\varepsilon)$ the matrix obtained from A by replacing zero in the left lower entry by $\varepsilon > 0$.

Suppose M is α-stable. Then, in particular, there is an eigenvector $y(\varepsilon)$ of $A(\varepsilon)$ such that with $M(\varepsilon) = \operatorname{span}\{y(\varepsilon)\}$ we have $\|P_M - P_{M(\varepsilon)}\| \leqslant K\|A - A(\varepsilon)\|^{1/\alpha} = K \cdot \varepsilon^{1/\alpha}$. Now

$$\|P_M - P_{M(\varepsilon)}\| \geqslant \frac{1}{\|y(\varepsilon)\|} \cdot \|P_M y(\varepsilon) - y(\varepsilon)\| > \frac{1}{2}\left(\sum_{j=1}^{n-1} \varepsilon^{2j/n}\right)^{1/2} > \tfrac{1}{2}\varepsilon^{1/n}.$$

So it follows that $\alpha \geqslant n$. From Theorem 3.1 one now sees that M in fact is strongly stable.

The rest of this section is devoted to producing an example of an α-stable subspace which is not strongly α-stable. Consider

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad M = \operatorname{span}\{e_1, e_2\}.$$

By Theorem 3.1, M is not strongly α-stable for any $\alpha < 4$. We shall show that it is α-stable for some $\alpha < 4$.

Consider a general perturbation A' of A. By applying a similarity transformation close to the identity we can always reduce to the case

$$A' = \begin{bmatrix} a & 1 & 0 & 0 \\ b & 0 & 1 & 0 \\ c & 0 & 0 & 1 \\ d & 0 & 0 & 0 \end{bmatrix}.$$

Actually, one can achieve this reduction by means of a similarity that depends analytically on the entries of A' (see [**A**]). The eigenvalues of A' are the solutions of the equation

$$\lambda^4 - a\lambda^3 - b\lambda^2 - c\lambda - d = 0. \tag{3.2}$$

In the sequel we shall use the fact that one can solve (3.2) explicitly. In fact, using Cardano's method we find two roots λ_1 and λ_2 of (3.2) such that $\lambda_1 + \lambda_2 = (a/2) - \alpha$ where $\alpha^2 = 2y + (a^2/4) + b$ and y is any solution to the cubic equation.

$$(ay - c)^2 = 4\left(2y + \frac{a^2}{4} + b\right)(y^2 + d) \tag{3.3}$$

(see, e.g., [**M**, pp. 671–676] or [**BS**, pp. 160–164]). Equation (3.3) can be rewritten in the form

$$y^3 + \tfrac{1}{2}by^2 + \frac{ac + 4d}{4}y + \tfrac{1}{8}\left((a^2 + 4b)d - c^2\right) = 0.$$

Substituting $w = y + (1/6)b$ leads to the equation $w^3 + pw = q = 0$ in w, where

$$p = \frac{ac + 4d}{4} - \frac{1}{12}b^2 \tag{3.4}$$

$$q = \frac{+b^3}{108} - \frac{b(ac + 4d)}{24} + \tfrac{1}{8}\big((a^2 + 4b)d - c^2\big). \tag{3.5}$$

Our first aim is to show that α, and hence also $\lambda_1 + \lambda_2$, is $0(\|A - A'\|^{1/3-\varepsilon})$ for arbitrarily small $\varepsilon > 0$. For this we need the following lemma.

LEMMA 3.3. *Let p and q be complex numbers depending polynomially on complex parameters $z = (z_1, \ldots, z_n)$ such that $|p(z)| \leqslant C_1\|z\|$, $|q(z)| \leqslant C_2\|z\|^2$ for $\|z\|$ small enough, where the positive constants C_1 and C_2 are independent of z. Given any $\delta > 0$ all solutions w of the cubic equation $w^3 + pw + q = 0$ will satisfy $|w| \leqslant C_3\|z\|^{2/3-\delta}$, for $\|z\|$ sufficiently small (and $C_3 > 0$ independent of z).*

Proof. Recall that w is found as follows. Put

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}, \qquad v = \sqrt[3]{-\left(\frac{q}{2}\right) - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

Then $w = u + v$, where the cubic roots in u and v are chosen such that $uv = -p/3$ (see e.g., [**M**] pp. 671–676 or [**BS**] pp. 160–164). We now consider two cases separately.

(1) $|p(z)| \geqslant C_4\|z\|^{4/3-\delta}$ for some $\delta > 0$.

(2) $|p(z)| \leqslant C_5\|z\|^{4/3-\delta}$ for some $\delta > 0$.

(We denote by C_j positive constants independent of z.)

In case (1), we write

$$\beta = \frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}.$$

Then

$$w = \sqrt[3]{-\beta} + \sqrt[3]{\beta - q} = \sqrt[3]{-\beta} + \sqrt[3]{\beta}\left(1 - \frac{1}{3}\frac{q}{\beta} - \frac{1}{9}\frac{q^2}{\beta^2} + \frac{5}{162}\frac{q^3}{\beta^3} + \cdots\right)$$

$$= \frac{1}{3}\sqrt[3]{\beta}\,\frac{q}{\beta} + \cdots.$$

Clearly this series converges, as $|q| < |\beta|$ for sufficiently small $\|z\|$. Indeed,

$$\frac{\beta}{q} = \frac{1}{2} - \sqrt{\frac{1}{4} + \frac{1}{27}\frac{p^3}{q^2}},$$

and

$$\left|\frac{p^3}{q^2}\right| \geqslant C_6 \cdot \|z\|^{4-3\delta-4} = C_6 \cdot \|z\|^{-3\delta} \to \infty \quad \text{as } \|z\| \to 0.$$

Thus, $\beta/q \to \infty$ as $\|z\| \to 0$, and consequently, $|q| < |\beta|$ for $\|z\|$ small enough. Now

for sufficiently small $\|z\|$ we have

$$|w| \leqslant C_7 \frac{|q|}{|\beta|^{2/3}}. \tag{3.6}$$

On the other hand,

$$\beta = \frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} = \frac{q}{2} - \frac{q}{2}\left(1 + \frac{1}{2} \cdot \frac{4}{27}\frac{p^3}{q^2} + \cdots\right) = -\frac{1}{27}\frac{p^3}{q} + \cdots.$$

So for $\|z\|$ sufficiently small

$$|\beta| = 0\left(\left|\frac{p^3}{q}\right|\right) \geqslant C_8\|z\|^{4-3\delta-2} = C_8\|z\|^{2-3\delta}.$$

So

$$|\beta^{2/3}| \geqslant C_q\|z\|^{4/3-2\delta},$$

and thus (using (3.6)), $|w| \leqslant C_{10} \cdot \|z\|^{2/3+2\delta}$. This finishes case (1).

In case (2) we estimate

$$|u|^3 \leqslant \left|\frac{q}{2}\right| + \sqrt{\left|\frac{q}{2}\right|^2 + \left|\frac{p}{3}\right|^3} \leqslant C_{12}\|z\|^{2-3/2\delta} \quad \text{so } |u| \leqslant C_{13}\|z\|^{2/3-1/2\delta}.$$

Likewise, also

$$|v| \leqslant C_{14} \cdot \|z\|^{2/3-1/2\delta}.$$

Thus,

$$|w| \leqslant C_{15}\|z\|^{2/3-1/2\delta}.$$

This proves case (2).

Returning to our original problem, we see that p and q given by (3.4) and (3.5) indeed satisfy the hypothesis of the Lemma 3.3. Thus, $|w| \leqslant C\|A - A'\|^{2/3-\delta}$ for $\|A - A'\|$ sufficiently small. Hence, the same holds for y, and thus, with $\alpha^2 = 2y + (a^2/4) + b$, we have $|\alpha| \leqslant C_1\|A - A'\|^{1/3-\varepsilon}$ for some arbitrarily small $\varepsilon > 0$ provided $\|A - A'\|$ is sufficiently small. Now, pick λ_1 and λ_2 two eigenvalues of A' such that

$$|\lambda_1 + \lambda_2| = \left|\frac{a}{2} - \alpha\right| \leqslant C\|A - A'\|^{1/3-\varepsilon}.$$

Let $x = (1, \chi_1, \chi_2, \chi_3)^T$ and $y = (1, y_1, y_2, y_3)^T$ be the corresponding eigenvectors; then

$$\begin{pmatrix} 1 \\ \chi_1 \\ \chi_2 \\ \chi_3 \end{pmatrix} = \begin{pmatrix} 1 \\ \lambda_1 \\ \lambda_1^2 \\ \lambda_1^3 \end{pmatrix} - a\begin{pmatrix} 0 \\ 1 \\ \lambda_1 \\ \lambda_1^2 \end{pmatrix} - b\begin{pmatrix} 0 \\ 0 \\ 1 \\ \lambda_1 \end{pmatrix} - c\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

and a similar formula holds for $(1, y_1, y_2, y_3)^T$ replacing λ_1 by λ_2. Assuming

$\lambda_1 \neq \lambda_2$, consider $(x - y)/(\lambda_1 - \lambda_2) = u$. We have:

$$u = \frac{x - y}{\lambda_1 - \lambda_2} = \begin{pmatrix} 0 \\ 1 \\ \lambda_1 + \lambda_2 \\ \lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2 \end{pmatrix} - a\begin{pmatrix} 0 \\ 0 \\ 1 \\ \lambda_1 + \lambda_2 \end{pmatrix} - b\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

So, if we put $M' = \operatorname{span}\{x, y\} = \operatorname{span}\{x, u\}$ we have $M' = \operatorname{Im}\begin{bmatrix} I \\ R \end{bmatrix}$, where

$$R = \begin{bmatrix} \lambda_1^2 - a\lambda_1 - b - \lambda_1(\lambda_1 + \lambda_2) + \lambda_1 a & \lambda_1 + \lambda_2 - a \\ \lambda_1^3 - a\lambda_1^2 - b\lambda_1 - c - \lambda_1(\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2 - a\lambda_1 - a\lambda_2 - b) & \lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2 - a(\lambda_1 + \lambda_2) - b \end{bmatrix}.$$

Now let

$$Q' = \begin{bmatrix} I & 0 \\ R & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}.$$

Then $Q'^2 = Q'$ is a projection onto M', while Q is a projection onto M. Clearly for $\|A - A'\|$ sufficiently small $\|Q' - Q\| = \|R\| \leq C\|A - A'\|^{1/3-\varepsilon}$, as the dominant term in R is in its upper right corner. Then (using (2.1)) we obtain

$$\|P_M - P_{M'}\| \leq \|Q - Q'\| \leq C\|A - A'\|^{1/3-\varepsilon}.$$

We have proved that for any positive ε there exists a constant $C > 0$ such that every 4×4 matrix A' with distinct eigenvalues has an invariant subspace M' such that

$$\|P_M - P_{M'}\| \leq C\|A - A'\|^{1/3-\varepsilon}.$$

This does not mean yet that M is $3(1 - 3\varepsilon)^{-1}$-stable, because we have to account for matrices A' with multiple eigenvalues. This can be done as follows:

Since the set of matrices with distinct eigenvalues is dense, given any 4×4 matrix A' there is a sequence of 4×4 matrices $\{A'_m\}_{m=1}^{\infty}$ with distinct eigenvalues such that $\lim_{m\to\infty} A'_m = A'$. By the already proved statement for every m there exists an A'_m-invariant subspace M'_m such that

$$\|P_M - P_{M'_m}\| \leq C\|A - A'_m\|^{1/3-\varepsilon}. \tag{3.7}$$

Now we make use of the fact that the set of all subspaces in $\mathbb{C}^n$ is a complete compact metric space in the metric given by $\|P_M - P_N\|$, where M, N are subspaces in $\mathbb{C}^n$. Moreover, a subspace N which is a limit (in this metric) of a sequence of subspaces $\{N_m\}_{m=1}^{\infty}$ can be described as

$$N = \Big\{\chi \in \mathbb{C} \mid \text{there is a sequence of vectors } \{\chi_m\}_{m=1}^{\infty} \text{ such that } \chi_m \in N_m \text{ for all } m \text{ and } \lim_{m\to\infty} \chi_m = \chi\Big\}.$$

A proof of this fact can be found, for example, in **[GLR1]** or **[GLR2]**. Returning to the sequence of subspaces $\{M'_m\}_{m=1}^{\infty}$ introduced above, choose a converging subsequence $\{M'_{m_k}\}_{k=1}^{\infty}$ with a limit M'. Using the description of a limit given by (3.8) we easily verify that M' is A'-invariant. Passing to the limit in (3.7) (with m replaced by m_k) we obtain the desired inequality

$$\|P_M - P_{M'}\| \leq C\|A - A'\|^{1/3-\varepsilon}.$$

We have proved that M is $3(1 - 3\varepsilon)^{-1}$-stable for every $\varepsilon > 0$.

Clearly, for ε small enough $3(1 - 3\varepsilon)^{-1} < 4$. This shows that although M is not strongly α-stable for any $\alpha < 4$, it is α-stable for every α with $3 < \alpha < 4$.

4. The stability index. It follows from Theorem 3.1 that every stable A-invariant subspace is strongly α-stable (and hence also α-stable) with α equal to the *height* of A. The height of A is defined as the size of the largest Jordan block in the Jordan canonical form corresponding to eigenvalues λ of A with $\dim \operatorname{Ker}(A - \lambda I) = 1$. (If for all eigenvalues λ of A one has $\dim \operatorname{Ker}(A - \lambda I) > 1$ we put the height of A equal to one.)

We next define the *stability index* of a stable A-invariant subspace M to be the infimum of all $\alpha > 0$ such that M is α-stable, and this number we denote by $s(M, A)$. To avoid trivialities we will apply this notion only to the non-trivial (i.e., different from (0) and the whole space) stable invariant subspaces.

In this section we study the stability index of a stable invariant subspace. Here too, many questions remain unsolved. We do give some bounds and estimates, which allow us to compute the stability index in special cases.

A standard reduction (see e.g., Section 15.3 in **[GLR]** or Section 8.1 in **[BGK]**) establishes the following fact.

PROPOSITION 4.1. *Let $\lambda_1, \ldots, \lambda_r$ be the distinct eigenvalues of the $n \times n$ matrix A such that $\dim \operatorname{Ker}(A - \lambda_i I) = 1$. Put $R_i = \operatorname{Ker}(\lambda_i I - A)^n$. Then for any stable $M \in \operatorname{Inv}(A)$ we have $s(M, A) = \max_{1 \leqslant i \leqslant r} s(M \cap R_i, A/R_i)$, where A/R_i stands for the restriction of A to its invariant subspace R_i.*

The next proposition gives upper and lower bounds for $s(M, A)$.

PROPOSITION 4.2. *$s(M, A)$ is less than or equal to the height of A, and $s(M, A) \geqslant 1$, provided $(0) \neq M \neq C^n$.*

Proof. The fact that $s(M, A)$ does not exceed the height of A is a consequence of the previous proposition and the remarks in the first paragraph of this section.

To show $s(M, A) \geqslant 1$, assume the contrary. By Proposition 4.1 we can assume without loss of generality that $A = J_n$ is the $n \times n$ Jordan block with zero eigenvalue and $M = \operatorname{span}\{e_1, \ldots, e_k\}$, $k < n$. Let $J_n(\varepsilon)$ be the matrix obtained from J_n by putting $\varepsilon > 0$ in the left lower corner. By assumption, there is a $J_n(\varepsilon)$-invariant subspace $M(\varepsilon)$ such that $\|P_M - P_{M(\varepsilon)}\| \leqslant C\|A - A(\varepsilon)\|^{1/\alpha} = C\varepsilon^{1/\alpha}$ for ε small enough. Let $\chi(\varepsilon) = (1, \varepsilon_i, \ldots, \varepsilon_i^{n-1})^T \in M(\varepsilon)$ be an eigenvector of $J_n(\varepsilon)$ corresponding to its eigenvalue ε_i which is an nth root of ε. Then, since $\chi(\varepsilon) \in M(\varepsilon)$, we have

$$\left\|(P_M - P_{M(\varepsilon)})\chi(\varepsilon)\right\| = \|P_M\chi(\varepsilon) - \chi(\varepsilon)\| \leqslant C\varepsilon^{1/\alpha}\|\chi(\varepsilon)\|.$$

Thus, in particular considering the $(k+1)$st coordinate of $P_M\chi(\varepsilon) - \chi(\varepsilon) = \varepsilon_i^k$ we have

$$|\varepsilon|^{k/n} = |\varepsilon_i|^k \leqslant C\varepsilon^{1/\alpha}\left(\sum_{j=0}^{n-1} |\varepsilon_i|^{2j}\right)^{1/2} \leqslant C\varepsilon^{1/\alpha}$$

for ε small enough, where $C' > 0$ is independent of ε. Clearly this is impossible for $\alpha < 1$.

Next we state another useful fact, which is easily verified.

PROPOSITION 4.3. *If* $M \in \mathrm{Inv}(A)$ *is stable then* $M^{\perp} \in \mathrm{Inv}(A^*)$ *is stable and* $s(M, A) = s(M^{\perp}, A^*)$.

By Proposition 4.1 the study of the stability index is reduced to the case where A is an $n \times n$ Jordan block with a single eigenvalue. Without loss of generality this single eigenvalue can be taken to be zero, as one sees by considering $A - \lambda I$ instead of A. So we can assume $A = J_n$. For $k = 1, \ldots, n-1$ let $s(k, n)$ be the stability index of the unique k-dimensional invariant subspace of J_n. (We ignore the trivial cases $k = 0$ and $k = n$.) From Proposition 4.3 it follows that $s(k, n) = s(n-k, n)$, and hence we can assume $k \leqslant (n/2)$ without loss of generality. In general, we are unable to find the exact value of $s(k, n)$. Inspection of the proof of Theorem 2.2 shows that $s(k, n) \geqslant n/k$, and that actually one can say $s(k, n) \geqslant n/j_0$ where $j_0 = \min\{j \mid p_j(\varepsilon, \ldots, \varepsilon_k) \neq 0$ for all choices of different $\varepsilon_1, \ldots, \varepsilon_k$ out of $\varepsilon_1, \ldots, \varepsilon_n\}$. However, these bounds are not very satisfying.

The next theorem is of more interest and is our main result. To state it we need the following definition. Given two positive integers k, n with $k < n$, we say that the pair (k, n) is *admissible* if no sum of k distinct nth roots of unity is equal to zero. In particular $(1, n)$ is admissible for all $n \geqslant 2$.

PROPOSITION 4.4. *If* n *is prime, then* (k, n) *is admissible for every* $1 \leqslant k < n$.

Proof. Use the fact that the cyclotomic polynomial $\chi^{n-1} + \chi^{n-2} + \cdots + 1$ is irreducible over the field of rational numbers (see, e.g., [**McD**]). Assume now that the primitive root $\zeta = \cos(2\pi/n) + i \sin(2\pi/n)$ satisfies an equation of type $\zeta^{m_1} + \zeta^{m_2} + \cdots + \zeta^{m_k} = 0$ $(n-1 \geqslant m_1 > m_2 > \cdots > m_k \geqslant 0$ are integers). Then the cyclotomic polynomial must be a divisor of $\chi^{m_1} + \chi^{m_2} + \cdots + \chi^{m_k}$ (see e.g., the proof of Theorem 38.1 in [**McD**]), which is possible only if $k = n$. This proves the proposition.

THEOREM 4.5. *Let* $k \leqslant n/2$. *Assume there is a smallest integer* p, $0 \leqslant p \leqslant k-1$, *such that* $(q, n-p)$ *is admissible for all integers* q *with* $k - p \leqslant q \leqslant k$. *Then* $s(k, n) \geqslant n - p$.

Proof. For ε small and positive, let $A(\varepsilon)$ be the matrix obtained from J_n by putting ε in the $(n-p, 1)$-entry in place of zero.

The eigenvalues of $A(\varepsilon)$ are zero and the $(n-p)$th roots of ε, call them $\eta_1, \ldots, \eta_{n-p}$. The eigenvectors corresponding to $\eta_1, \ldots, \eta_{n-p}$ are

$$y_1 = \left(1, \eta_1, \ldots, \eta_1^{n-p-1}, 0, \ldots, 0\right)^T, \ldots, y_{n-p} = \left(1, \eta_{n-p}, \ldots, \eta_{n-p}^{n-p-1}, 0, \ldots, 0\right)^T$$

and there is a Jordan chain corresponding to the eigenvalue 0:

$$e_1 - \varepsilon e_{n-p+1}, e_2 - \varepsilon e_{n-p+2}, \ldots, e_p - \varepsilon e_n,$$

where by e_q we denote the unit coordinate vector with 1 in the qth position and zeros elsewhere. Consequently, every k-dimensional $A(\varepsilon)$-invariant subspace has the form

$$N = \mathrm{span}\{y_{i_1}, \ldots, y_{i_q}, e_1 - \varepsilon e_{n-p+1}, e_2 - \varepsilon e_{n-p+2}, \ldots, e_{k-q} - \varepsilon e_{n-p+(k-q)}\}$$

for some $q (k - p \leqslant q \leqslant n - p; q \leqslant k)$ and some choice of the vectors $y_{i_1}, \ldots, y_{i_q}$ among $y_1, \ldots, y_{n-p}$. Introduce also the subspace $M = \mathrm{span}\{y_{i_1}, \ldots, y_{i_q},$

$e_1, e_2, \ldots, e_{k-q}\}$. Clearly,

$$\|P_N - P_M\| \leqslant K_1\varepsilon \tag{4.1}$$

for some constant K_1 independent of ε. After some elementary transformations of the set of spanning vectors for M, we obtain that M is spanned by $e_1, e_2, \ldots, e_{k-q}$ and by the vectors

$$f_t := e_{k-q+1} + \eta_{i_t} e_{k-q+2} + \cdots + \eta_{i_t}^{s-1} e_{n-p}, \qquad t = 1, \ldots, q$$

where $s = n - p - (k - q)$. (It follows from our assumptions on k and p that s is positive.) Denoting by M_0 the k-dimensional J_n-invariant subspace, it is easily seen now that

$$\|P_{M_0} - P_M\| = \|P_{\operatorname{span}\{f_1, \ldots, f_q\}} - P_{\operatorname{span}\{e_{k-q+1}, \ldots, e_k\}}\|.$$

Consider the $n \times q$ matrix W with columns $f_1, \ldots, f_q$: $W = [f_1, f_2, \ldots, f_q]$, and observe that the rows $k - q + 1$ through k in W form the Vandermonde matrix

$$W_0 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \eta_{i_1} & \eta_{i_2} & \cdots & \eta_{i_q} \\ \vdots & \vdots & & \vdots \\ \eta_{i_1}^{q-1} & \eta_{i_2}^{q-1} & \cdots & \eta_{i_q}^{q-1} \end{bmatrix}, \tag{4.2}$$

which is invertible. A computation (see e.g., [**GLR**, Section 16.5]) shows that the matrix WW_0^{-1} has the $(k + 1, q)$-entry equal to $\eta_{i_1} + \cdots + \eta_{i_q}$.

Now use the assumption that $(q, n - p)$ is an admissible pair (since $k - p \leqslant q \leqslant k$). It follows that there is a constant $K > 0$ such that $|\lambda_1 + \cdots + \lambda_q| \geqslant K$ for any choice of q $(n - p)$th roots of unity $\lambda_1, \ldots, \lambda_q$. Then also

$$|\eta_{i_1} + \cdots + \eta_{i_q}| \geqslant K \cdot |\varepsilon|^{1/(n-p)} \tag{4.3}$$

for any choice of the q numbers $\eta_{i_1}, \ldots, \eta_{i_q}$. Now consider the last column in WW_0^{-1}. Call this vector χ. Clearly $\chi \in \operatorname{span}\{f_1, \ldots, f_q\}$ and χ is of the form $(0, \ldots, 0, 1, \eta_{i_1} + \cdots + \eta_{i_q}, \ldots)^T$ where the first $k - 1$ entries are zero. Then (4.2) gives

$$\left\| P_{M_0} - P_M \right\| \geqslant \frac{1}{\|\chi\|} \cdot \left\| P_{\operatorname{span}\{f_1, \ldots, f_q\}}\chi - P_{\operatorname{span}\{e_{k-q+1}, \ldots, e_k\}}\chi \right\|$$

$$= \frac{1}{\|\chi\|}\|\chi - e_k\| \geqslant \frac{1}{\|\chi\|} \cdot |\eta_{i_1} + \cdots + \eta_{i_q}|.$$

For ε small enough we obtain by using (4.3) that

$$\|P_{M_0} - P_M\| \geqslant K_2|\varepsilon|^{1/(n-p)}$$

for some constant $K_2 > 0$ independent of ε. Taking into account (4.1), we get

$$\|P_N - P_{M_0}\| \geqslant K_3|\varepsilon|^{1/(n-p)} \tag{4.4}$$

for any k-dimensional $N \in \operatorname{Inv}(A(\varepsilon))$, where $K_3 > 0$ is independent of both ε and N. Inequality (4.4) shows that $M_0 \in \operatorname{Inv}(J_n)$ cannot be α-stable for $\alpha < n - p$, which proves $s(n, k) \geqslant n - p$, and our theorem is proved.

As a corollary to Theorem 4.5 and Proposition 4.4 we have

COROLLARY 4.6. *Let r be the smallest number such that $n - r$ is prime. Then $s(k, n) \geqslant n - r$ for all $k (1 \leqslant k \leqslant n - 1)$. In particular, $s(k, n) = n$ for prime n.*

It follows, for example, that $s(2, 4) \geqslant 3$, and combining with our results obtained in the example in the previous section we see that $s(2, 4) = 3$.

In terms of the stability index we can say precisely when α-stability for a subspace M implies strong α-stability. To state the result we need the following notion. Given an $n \times n$ matrix A and $M \in \mathrm{Inv}(A)$ we introduce the *height of A with respect to M* as follows. Let $\{\lambda_1, \ldots, \lambda_r\}$ be the set of all distinct eigenvalues of A with $\dim \mathrm{Ker}(\lambda_i I - A) = 1$, and let $\sigma = \{i | 1 \leqslant i \leqslant r, 0 \neq M \cap \mathrm{Ker}(\lambda_i I - A)^n \neq \mathrm{Ker}(\lambda_i I - A)^n\}$. Then the height of A with respect to M, denoted by $h(M, A)$ is the height of $A|_N$ where N is the sum of the subspaces $\mathrm{Ker}(\lambda_i I - A)^n$, $i \in \sigma$. Again, the height of A with respect to M is defined to be 1 if σ is empty. With this definition we have the following result.

THEOREM 4.7. *The following two statements are equivalent for a stable A-invariant subspace M:*

(a) $s(M, A) = h(M, A)$

(b) *For any α: M is strongly α-stable if and only if M is α-stable.*

Proof. The case when $\sigma = \phi$ is easy. In this case M is the sum of root subspaces and hence Lipschitz stable and strongly Lipschitz stable by Theorem 2.2, hence Theorem 4.7 follows from the definition of $h(M, A)$. So we assume $\sigma \neq \phi$ from now on. Clearly, both $s(M, A)$ and $h(M, A)$ as well as α-stability and strong α-stability for M depend only on $M \cap N$ and on the restriction $A|_N$, where N is as above. So we can assume $A = J_1 \oplus \cdots \oplus J_r$, $M = M_1 \oplus \cdots \oplus M_r$ where M_i is J_i-invariant with $0 \neq M_i \neq \mathrm{Ker}(\lambda_i I - A)^{n_i}$. We also can arrange it so that $n_1 \geqslant n_2 \geqslant \cdots \geqslant n_r$, in which case $h(M, A) = n_1$.

Further, M is α-stable or strongly α-stable with respect to A if and only if each M_i is α-stable or strongly α-stable with respect to J_i.

Now suppose (a) holds. Then if M is α-stable α must be larger than or equal to n_1. Then certainly $\alpha \geqslant n_i$ for $1 \leqslant i \leqslant r$. By Theorem 3.1 it then follows that M is strongly α-stable. As strong α-stability implies α-stability always, this proves (b).

Conversely, suppose (b) holds for any α. Then, if M is α-stable, it must be strongly α-stable and hence Theorem 3.1 implies that $\alpha \geqslant n_i$ for each n_i. Thus, $\alpha \geqslant h(M, A)$. This shows $s(M, A) \geqslant h(M, A)$. But clearly M is $h(M, A)$-(strongly) stable. So $s(M, A) = h(M, A)$. Note that we use here a sharpening of Proposition 4.2. The theorem is proved.

We have introduced the notion of stability index and have produced some bounds on the stability index. These bounds allowed us to find the precise value of the stability index in some cases. It is an open problem to find the stability index in the general case.

5. Open problems. We summarize here some open problems (including those mentioned already in the text) that arise naturally in the context of α-stability.

Problem 5. For a given α, describe (in terms of the intersections of M with the root subspaces $\mathrm{Ker}(A - \lambda I)^n$) all α-stable subspaces M of a matrix A.

Problem 5.2. Describe the relationship between α-stability and strong α-stability. In particular, a more concrete problem is: When does α-stability imply strong α-stability? Again, the answer should be given in terms of intersections of the invariant subspace in question with the root subspaces.

Problem 5.3. Find the stability index of every stable A-invariant subspace M, in terms of the dimensions of $\operatorname{Ker}(A - \lambda I)^n$ and $M \cap \operatorname{Ker}(A - \lambda I)^n$, for every eigenvalue λ of A. In particular, is the stability index always a positive integer?

The solution of Problem 5.3 hinges upon solution to the following:

Problem 5.4. Find $s(k, n)$ for all positive integers k and n, $1 \leqslant k \leqslant n - 1$.

Again, partial answers and bounds for $s(k, n)$ are found in the text. It follows, in particular, that $s(k, n) = n$ for all (k, n) with $1 \leqslant k \leqslant n - 1$; $2 \leqslant n \leqslant 5$; except that $s(2, 4) = 3$. Also, $s(k, 6) = 6$ if $k = 1, 5$ and $5 \leqslant s(k, 6) \leqslant 6$ for $k = 2, 3, 4$. Hence the first concrete open question is:

Problem 5.5. Find the values of $s(2, 6) = s(4, 6)$ and $s(3, 6)$.

Acknowledgments. We are grateful to Prof. M. A. Kaashoek for his interest in the material of this paper and for his suggestion to write it up.

REFERENCES

[A] Arnold, V. I.: On matrices depending on parameters, *Uspehi Math. Nauk*, 26 (1971) 101–114 [Russian].

[BGK] Bart, H.; Gohberg, I. and Kaashoek, M. A.: Minimal factorizations of matrix and operator functions, OT1, Birkhäuser Verlag, Basel, 1979.

[BS] Bronshtein, I. N. and Semendyayev, K. A.: A guide-book to mathematics for technologists and engineers, MacMillan, New York, 1964.

[CD] Campbell, S. and Daughtry, J.: The stable solutions of quadratic matrix equations, *Proc. AMS* 74 (1979), 19–23.

[GK] Gohberg, I. C. and Krein, M. G.: Introduction to the theory of linear nonselfadjoint operators, Translations of Mathematical Monographs, Vol. 24, American Mathematical Society, Providence, R.I., 1970; reprinted 1988.

[GLR1] Gohberg, I., Lancaster, P. and Rodman, L.: Matrix Polynomials, Academic Press, 1982.

[GLR2] Gohberg, I., Lancaster, P. and Rodman, L.: Invariant Subspaces of Matrices with Applications, J. Wiley & Sons, New York, etc., 1986.

[GR] Gohberg, I. and Rodman, L.: On distance between lattices of invariant subspaces of matrices, *Linear Algebra Appl.*, 76 (1986) 85–120.

[GvL] Golub, G. H. and Van Loan, C. F.: Matrix Computations, Johns Hopkins University Press, Baltimore, 1983.

[KvdMR] Kaashoek, M. A., van der Mee, C. V. M. and Rodman, L.: Analytic operator functions with compact spectrum II: Spectral pairs and factorization, *Integral Equations and Operator Theory*, 5 (1982) 791–827.

[McD] MacDuffee, C. C.: An Introduction to Abstract Algebra, J. Wiley & Sons, New York, 1940.

[M] Meyers Handbuch uber die Mathematik, Bibliographisches Institut, Mannheim, Wien, Zurich, 1972. [German].

[RR] Ran, A. C. M. and Roozemond, L.: On strong α-stability of invariant subspaces of matrices, in: The Gohberg anniversary collection, OT40, Birkhäuser Verlag, Basel, 1989, 427–435.

[W] Wilkinson, J. H.: The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1965.

LETTERS TO THE EDITOR

To the Editor:

In a recent note [1], Friedberg gives counterexamples to the principal axis theorem ("a real symmetric matrix is orthogonally equivalent to a diagonal matrix") over some fields other than the reals, and he concludes by asking for which fields the theorem is true. The answer [2] is that it is true precisely when the field is real closed or an intersection of real closed fields. Furthermore, the principal axis theorem is actually true if we make only the seemingly weaker assumption that every symmetric matrix over the field is similar to a diagonal matrix. The fields in [2] are assumed to have characteristics different from 2, but it is easy to see that the theorem can never hold in characteristic 2: any 2×2 symmetric matrix there with all entries equal is nilpotent and thus cannot be diagonalized.

REFERENCES

1. S. H. Friedberg, Extending the principal axis theorem to fields other than $\mathbb{R}$, *Amer. Math. Monthly*, 97 (1990) 147–149.
2. W. C. Waterhouse, Self-adjoint operators and formally real fields, *Duke Math. J.*, 43 (1976) 237–243.

William C. Waterhouse
Penn State
University Park, PA 16802

To the Editor:

In his paper (this MONTHLY, vol. 95, pp. 950–952), L. J. Gerston introduced a new algorithm of computing the rank of a matrix. In fact, this method had been published in 1978, by Professor Xie Bangjie of Jilin University [1]. In the book he referred to it as depression of order.

REFERENCE

1. Xie Bangjie, Linear algebra, People's Education Press, 1978.

Xu Pingya
Nanjing College of Electric Power
Jiangsu, China

To the Editor:

It was good to see the note by T. W. Hungerford in the MONTHLY (January, 1990). As he said, some presentations of Galois theory consider field extensions of the rationals given by radicals and claim that they will be Galois when appropriate roots of unity are adjoined. He gave an example to show that this does not have to be true with the "obvious" choices for appropriate roots of unity. I want to reinforce his warning by offering another example. Take $\mathbb{Q}(i)$ (with $i^2 = -1$), and let $K = \mathbb{Q}(i)(u)$ with $u^3 = 2 + i$. Clearly $2 + i$ is not a cube in $\mathbb{Q}(i)$, since $(2 + i)\overline{(2 + i)} = 5$ is not a cube in $\mathbb{Q}$; thus K has degree 6 over $\mathbb{Q}$. I claim that it will not become Galois no matter what roots of unity are adjoined.

To see this, we first need to know a little about L, the smallest Galois extension of $\mathbb{Q}$ containing K. Clearly L must contain a cube root of 1, say ζ. The complex

conjugation on $\mathbb{Q}(i)$ extends to an automorphism of L, which will take u to a cube root v of $2 - i$. Hence L contains a cube root of 5, and the Galois extension $M = \mathbb{Q}(\zeta, \sqrt[3]{5})$ is inside L. It is easy to see that L is the composite $K \cdot M$. A little more work shows further that L has degree 36 over $\mathbb{Q}$, and hence $K \cap M = \mathbb{Q}$. I omit the proof; elementary ones are possible, though it's faster to use the fact that $2 + i$ and $2 - i$ are relatively prime elements in $\mathbb{Z}[i]$.

THEOREM. *Let E be any extension of $\mathbb{Q}$ with abelian Galois group. Then $K \cdot E$ does not contain L.*

Proof. If $K \cdot E$ contains L, then of course $K \cdot E = L \cdot E$; hence the field $L \cdot E$ is the composite of two Galois extensions and is Galois over $\mathbb{Q}$. Let G be the Galois group Gal $(L \cdot E/\mathbb{Q})$. Let N_E be the normal subgroup fixing E, and define N_L and N_M similarly. Let H be the subgroup (not normal) that fixes K. Since $K \cap M = \mathbb{Q}$, we have $H \cdot N_M = G$. Hence H maps homomorphically onto the quotient $G/N_M \cong \text{Gal}(M/\mathbb{Q}) \cong S_3$. However, since $K \cdot E = L \cdot E$, we also have $H \cap N_E = \{1\}$, and hence H must map one-to-one into the abelian group $G/N_E \cong \text{Gal}(E/\mathbb{Q})$. This contradiction proves the theorem. □

The theorem yields what we want, since roots of unity generate abelian extensions of $\mathbb{Q}$. The example also illustrates exactly what is wrong in the erroneous attempted proofs: if we have an extension of $\mathbb{Q}$ containing elements (like $2 + i$ and $2 - i$) conjugate over $\mathbb{Q}$, their roots may well give extensions differing by more than roots of unity.

William C. Waterhouse
Penn State
University Park, PA 16802

To the Editor:

Dr. Larry Zalcman has pointed out some errors in my "Putnam Trivia" article (96 (1989), 711–713). The answer to question 11 should have been: Arthur Rubin and D. J. Newman. The answer to question 23 should have been: City College of New York, Stanford, and the University of Maryland. The answer to question 17 should include a second place finish by Dartmouth in 1962, since *at that time* they were strictly a four year liberal arts college. I regret having made these errors.

Joe Gallian

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

A New Derivation of Stirling's Approximation to $n!$

GEORGE MARSAGLIA
Supercomputer Computations Research Institute and Department of Statistics, Florida State University

JOHN C. W. MARSAGLIA
Department of Computer Science, Western Oregon State College, Monmouth, OR

Stirling's formula, $n! \sim n^n e^{-n}\sqrt{2\pi n}$, perhaps the most widely used relation in asymptotic analysis, is usually mentioned in Calculus II courses. It is the first term in an asymptotic series whose derivation requires deeper methods than are ordinarily covered in Calculus II. It would be nice to have a development that only requires methods within the scope of the usual Calculus II course. That is the purpose of this note.

The classical method for developing Stirling's approximation uses Euler's summation formula:

$$f(1) + f(2) + \cdots + f(n) = \int_1^n f(x) + \sum_{k=1}^{m} \frac{B_k}{k!}\left(f^{k-1}(n) - f^{k-1}(1)\right) + R_m$$

with the B's Bernoulli numbers and $f(x) = \ln(x)$. This leads to

$$\ln(n!) = \ln(1) + \cdots + \ln(n) = \left(n + \frac{1}{2}\right)\ln(n) - n + c + \frac{1}{12n} - \frac{1}{360n^3} + \cdots .$$

The development is somewhat delicate. It requires establishing that the constant c is $\ln\sqrt{2\pi}$ (missing from Stirling's original development). Then the asymptotic series for $\ln(n!)$ with the Bernoulli numbers must be exponentiated to get the asymptotic series for $n!$. See, for example, Knopp [**1**, Chap. 14].

We will derive an asymptotic series for $n!$ directly, without recourse to Bernoulli numbers and Euler's summation formula for $\ln(n!)$. No new series will appear, of course, since a function can have at most one asymptotic series, but the derivation appears more direct. It involves only elementary transformations of integrals and term-by-term integration of a series.

Write

$$n! = \int_0^\infty x^n e^{-x}\,dx = n^n e^{-n}\int_0^\infty \left(\frac{x}{n}e^{1-x/n}\right)^n dx.$$

Then make a change of variable, $y = x/n$, to give

$$n! = n^{n+1}e^{-n}\int_0^\infty \left(ye^{1-y}\right)^n dy,$$

which may be written

$$n! = n^{n+1}e^{-n}\left[\int_0^1 \left(ue^{1-u}\right)^n du + \int_1^\infty \left(ve^{1-v}\right)^n dv\right]. \tag{1}$$

The idea, after Watson [**3**] and Marsaglia [**2**], is to break the function ye^{1-y} into two monotone parts, ue^{1-u}, increasing for $0 \leqslant u \leqslant 1$, and ve^{1-v}, decreasing for

$1 \leqslant v < \infty$. Consider the transformation defined by the relation

$$ve^{1-v} = e^{-\frac{1}{2}z^2}.$$

This defines a one-to-one relation between v in $[1, \infty)$ and z in $[0, \infty)$. We cannot express v explicitly in terms of z, but we may assume a power series,

$$v = 1 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 + \cdots,$$

then find the coefficients by this device: Let $E(z)$ represent $e^{-\frac{1}{2}z^2}$. Differentiating $ve^{1-v} = E(z)$ with respect to z gets $v'E(z)/v - v'E(z) = -zE(z)$ and hence $v'(v-1) = zv$. Multiplying the resulting series and equating like powers of z establishes that $a_1^2 = 1$, $a_2 = 1/3$, $a_3 = 1/36a_1$ and, for $n \geqslant 3$,

$$(n+1)a_1 a_n = a_{n-1} - 2a_2 a_{n-1} - 3a_3 a_{n-2} - \cdots - (n-1)a_{n-1}a_2. \quad (3)$$

There are two choices for a_1: 1 or -1. Since v increases with z, $a_1 = 1$ is the right choice. This gives $a_2 = 1/3$, $a_3 = 1/36$. Then $a_4, a_5, \ldots$ are completely determined by the above recursion. The coefficients $a_1, a_2, \ldots, a_{11}$ are

$$1, \frac{1}{3}, \frac{1}{36}, \frac{-1}{270}, \frac{1}{4320}, \frac{1}{17010}, \frac{-139}{5443200}, \frac{1}{204120}, \frac{-571}{2351462400}, \frac{-281}{1515591000}, \frac{163879}{2172751257600}.$$

On the other hand, the choice $a_1 = -1$ leads to the same numerical values for subsequent a's, except that it changes the sign at the odd indices. But this choice provides a series for the u defined by $ue^{1-u} = e^{-\frac{1}{2}z^2}$ with u in $[0, 1]$. In summary: if $a_1 = 1$ and $a_2, a_3, \ldots$ are determined as above, then

$$v = 1 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 + a_5 z^5 + \cdots$$

is the solution to $ve^{1-v} = e^{-\frac{1}{2}z^2}$, with v in $[1, \infty)$, while

$$u = 1 - a_1 z + a_2 z^2 - a_3 z^3 + a_4 z^4 - a_5 z^5 + \cdots$$

is the solution to $ue^{1-u} = e^{-\frac{1}{2}z^2}$ with u in $[0, 1]$.

We are now ready to make the transformation of integrals in (1). Let

$$u = 1 - a_1 z + a_2 z^2 - a_3 z^3 + \cdots \qquad du/dz = -a_1 + 2a_2 z - 3a_3 z^2 + \cdots,$$

with boundary conditions

$$z = \infty \quad \text{when} \quad u = 0; \qquad z = 0 \quad \text{when} \quad u = 1.$$

This transforms the left integral in (1), (after changing dz to $-dz$ and switching limits), to

$$\int_0^1 [ue^{1-u}]^n \, du = \int_0^\infty e^{-\frac{1}{2}nz^2}\left(a_1 - 2a_2 z + 3a_3 z^2 - 4a_4 z^3 + \cdots\right) dz.$$

To transform the second integral in (1), let

$$v = 1 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots, \qquad dv/dz = a_1 + 2a_2 z + 3a_3 z^2 + \cdots,$$

with boundary conditions

$$z = 0 \quad \text{when} \quad v = 1; \qquad z = \infty \quad \text{when} \quad v = \infty.$$

This transforms the second integral to

$$\int_1^\infty [ve^{1-v}]^n\,dv = \int_0^\infty e^{-\frac{1}{2}nz^2}\left(a_1 + 2a_2z + 3a_3z^2 + 4a_4z^3 + \cdots\right)dz.$$

The odd powers of z cancel on adding these two integrals. Then term-by-term integration, using

$$2\int_0^\infty e^{-\frac{1}{2}nz^2}z^{2k}\,dz = \left(\frac{2}{n}\right)^{k+\frac{1}{2}}\Gamma\left(k+\frac{1}{2}\right)$$

yields, from (1),

$$n! = n^{n+1}e^{-n}\left[\left(\frac{2}{n}\right)^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right) + 3a_3\left(\frac{2}{n}\right)^{\frac{3}{2}}\Gamma\left(\frac{3}{2}\right) + 5a_5\left(\frac{2}{n}\right)^{\frac{5}{2}}\Gamma\left(\frac{5}{2}\right) + \cdots\right].$$

Finally, factoring out the first term, with $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2})$, etc., gives the desired expression for $n!$:

$$n! = n^ne^{-n}\sqrt{2\pi n}\left(1 + \frac{3a_3}{n} + \frac{3\cdot 5a_5}{n^2} + \frac{3\cdot 5\cdot 7a_7}{n^3} + \frac{3\cdot 5\cdot 7\cdot 9a_9}{n^4} + \cdots\right),$$

with the a's determined by $a_1 = 1$, $a_2 = 1/3$ and recursion (3).

If we write this asymptotic expansion in the form

$$n! = n^ne^{-n}\sqrt{2\pi n}\left(1 + \frac{c_1}{n} + \frac{c_2}{n^2} + \frac{c_3}{n^3} + \frac{c_4}{n^4} + \cdots\right),$$

then coefficients $c_1, c_2, \ldots, c_9$ are:

$$\frac{1}{12}, \frac{1}{288}, \frac{-139}{51840}, \frac{-571}{2488320}, \frac{163879}{209018880}, \frac{5246819}{75246796800},$$
$$\frac{-534703531}{902961561600}, \frac{-4483131259}{86684309913600}, \frac{432261921612371}{514904800886784000}.$$

The sign pattern is $++--$.

In the classroom, students would be reminded that the above is an asymptotic series. It does not converge, as do most of the series they have encountered, but it still provides valuable information about $n!$. Successive partial sums approach the true value, $n!$, then diverge. The approach is erratic. Improvement is not monotonic, nor do successive sums, or pairs of sums, bracket the target value.

Multiple-precision computer routines are necessary for demonstrating the approach to $n!$ before divergence, however, unless n is small. For example, 69! is an integer of 99 digits, the limiting size for scientific notation in many hand calculators. It begins 17112245... and ends ...4224000000000000000. Let $S_0, S_1, \ldots$ be successive partial sums for Stirling's series, so that $S_0 = n^ne^{-n}\sqrt{2\pi n}$, $S_1 = n^ne^{-n}\sqrt{2\pi n}\,(1 + \frac{1}{12}n)$, etc. Then for $n = 69$, S_0 gives 0.1709159×10^{99}, accurate to some 3 places, while S_1 provides about 6 places. Going much further, S_{25} provides 48 places of accuracy, S_{50} provides 70, S_{75} provides 94 and S_{100} provides 111 places of accuracy: $S_{100} = 69! - 0.22 \times 10^{-10}$, while $S_{101} = 69! + \varepsilon$ with $0 < \varepsilon < 10^{-20}$.

The erratic behavior, then divergence, of the series may be demonstrated for small values of n, which need not be an integer, of course, but only for integers do

we know $n!$ exactly. Consider $n = 4$. Then $4! = 24$, $S_0 = 23.506\ldots$. Next, $S_1 = 23.995$ gives about 4 places of accuracy, and so on to the best sum, $S_{25} = 24.00000000002184\ldots$. After that, the sums slowly get less accurate, with $S_{65} = 24.079$, $S_{70} = 28.026$, $S_{71} = -6.281$, $S_{74} = 276.44$, $S_{75} = -2098.1$, and then chaos.

Knopp [1] is a good source for properties and examples of asymptotic series, §65–66, pages 535–555.

REFERENCES

1. K. Knopp, Theory and Application of Infinite Series, Blackie, Glasgow, 1951.
2. J. C. W. Marsaglia, The incomplete gamma function and Ramanujan's rational approximation to e^x. *Journal of Statistical Computation and Simulation*, 24 (1986) 163–169.
3. G. N. Watson, Theorems stated by Ramanujan (v): Approximations connected with e^x, *Proc. London Math. Society*, 29 (1929) 293–308.

Fibonacci Numbers and Aitken Sequences Revisited

M. J. Jamieson
Department of Computing Science, University of Glasgow, Glasgow, Scotland

Phillips [1] has shown that Aitken's transformation applied to the sequence of ratios of consecutive Fibonacci numbers yields further ratios of Fibonacci numbers and that this remains true if the transformation is applied repeatedly to the new sequences so obtained. He has shown also that the secant and Newton Raphson methods for solving the quadratic equation

$$x^2 - x - 1 = 0 \tag{1}$$

satisfied by the sequence limit (the golden ratio τ), yield estimates which are ratios of consecutive Fibonacci numbers if the starting values (two for the secant method, one for Newton Raphson) are such ratios. The purpose of this brief note is to generalize these results to ratios of every pth Fibonacci number.

The Fibonacci numbers are given by

$$u_{i+1} = u_i + u_{i-1}, \qquad i > 1 \tag{2}$$

with $u_1 = 1$ and $u_2 = 1$. (The initial subscripts are chosen here to make subsequent notation simpler.) There are p different possible sequences of ratios of every pth Fibonacci number depending on the choice of denominator from $u_1, u_2, \ldots, u_p$ for the first one. These sequences, $\{w_i^{(k)}\}$, are given by

$$w_i^{(k)} = u_{p(i+1)-k}/u_{pi-k}, \qquad i \geq 1 \tag{3}$$

where k in equation (3) and subsequently in this note lies in the range $0, 1, \ldots, (p-1)$. In Phillips's sequence p has the value unity. The limit of the sequence of ratios of consecutive Fibonacci numbers satisfies the quadratic equation (1). We wish now to find a quadratic equation to be satisfied by the limit of the sequence $\{w_i^{(k)}\}$. It is well known that

$$u_i = \left[\tau^i - (-\tau)^{-i}\right]/\sqrt{5}\,. \tag{4}$$

From equation (2) we may prove by induction that

$$u_{i+p} = v_p u_i - (-1)^p u_{i-p}, \qquad i > p \tag{5}$$

where v_p depends only on p, being the pth term of the series of Lucas numbers $1, 3, 4, \ldots$ which satisfy the same recurrence relation (2) as the Fibonacci numbers but have different starting values. Dividing equation (5) through by u_i and replacing i by $ip - k$ we find

$$w_i^{(k)} = v_p - (-1)^p / w_{i-1}^{(k)}, \qquad i > 1. \tag{6}$$

The limit of the sequence $\{w_i^{(k)}\}$ as i tends to infinity therefore satisfies the quadratic equation

$$x^2 - v_p x + (-1)^p = 0. \tag{7}$$

From equations (3) and (4) we see that this limit is τ^p which is independent of k. From the properties of the roots of a quadratic or from the properties of Lucas numbers we can determine v_p explicitly and replace equation (7) by

$$x^2 - \left[\tau^p + (-\tau)^{-p}\right] x + (-1)^p = 0. \tag{8}$$

Let us consider the secant method approximation to a root of equation (8). We construct a new estimate to the root from two existing estimates. Let these be drawn from all the possible sequences $\{w_i^{(k)}\}$; they could be, for example, $w_i^{(k)}$ and $w_j^{(l)}$ in which case the secant estimate is

$$\begin{aligned} \left[w_i^{(k)} w_j^{(l)} - (-1)^p\right] / \left[w_i^{(k)} + w_j^{(l)} - \tau^p - (-\tau)^{-p}\right] &= w_{i+j}^{(k+l)} \qquad \text{if } k + l < p \\ &= w_{i+j-1}^{(k+l-p)} \quad \text{if } k + l \geq p \end{aligned} \tag{9}$$

where l is in the range $0, 1, \ldots, (p-1)$. Equation (9) is obtained from equations (3) and (4). Hence we always get another ratio of every pth Fibonacci number. We note that only with $k = l = 0$ do we find that $w_i^{(k)}$, $w_j^{(l)}$ and the secant estimate belong to the same sequence ($\{w_i^{(0)}\}$ in this case).

We now consider the Newton Raphson approximation to a root of equation (8). We construct a new estimate to the root from $w_i^{(k)}$. The formula is the same as the left-hand side of equation (9) with $w_j^{(l)}$ replaced by $w_i^{(k)}$, and again we obtain another ratio of every pth Fibonacci number; it belongs to the same sequence as $w_i^{(k)}$ only for $k = 0$.

Finally we apply Aitken's acceleration formula to $w_{i+m}^{(k)}$, $w_i^{(k)}$ and $w_{i-m}^{(k)}$. From equation (9) we have

$$\begin{aligned} w_i^{(k)} w_j^{(k)} &= (-1)^p + \left[w_i^{(k)} + w_j^{(k)} - \tau^p - (-\tau)^{-p}\right] w_{i+j}^{(2k)} \qquad \text{if } 2k < p \\ &= (-1)^p + \left[w_i^{(k)} + w_j^{(k)} - \tau^p - (-\tau)^{-p}\right] w_{i+j-1}^{(2k-p)} \quad \text{if } 2k \geq p. \end{aligned} \tag{10}$$

The Aitken transformation is

$$\begin{aligned} &\left[w_{i+m}^{(k)} w_{i-m}^{(k)} - w_i^{(k)} w_i^{(k)}\right] / \left[w_{i+m}^{(k)} - 2w_i^{(k)} + w_{i-m}^{(k)}\right] \\ &\quad = w_{2i}^{(2k)} \qquad \text{if } 2k < p \\ &\quad = w_{2i-1}^{(2k-p)} \quad \text{if } 2k \geq p \end{aligned} \tag{11}$$

where equation (11) is obtained from equation (10) by first substituting $i + m$ for i and $i - m$ for j and then i for i and i for j. We obtain a ratio of every pth Fibonacci number which belongs to the same sequence as the estimates put into

Aitken's transformation only for $k = 0$. With m set to unity we see that the transformation applied to members of the sequence $\{w_i^{(0)}\}$ produces further members of that sequence, namely the subsequence $\{w_{2i}^{(0)}\}$. By replacing i by $2i, 4i, 8i, \ldots$ and m by $2, 4, 8, \ldots$ in equation (11) we see that subsequent applications of Aitken's transformation to the new subsequences so obtained yields the subsequences $\{w_{4i}^{(0)}\}, \{w_{8i}^{(0)}\}, \{w_{16i}^{(0)}\}, \ldots$.

In conclusion we see that Phillips's results can be generalized to ratios of every pth Fibonacci number. There are p different sequences of such ratios and the application of the secant, Newton Raphson and Aitken acceleration methods return results *in the same sequence* only for one of these, namely the sequence $u_{2p}/u_p, u_{3p}/u_{2p}, \ldots$.

The conclusions are illustrated below for $p = 2$ in which case v_p of equation (7) has the value 3.

TABLE. The sequences $\{w_i^{(0)}\}$ and $\{w_i^{(1)}\}$ for $p = 2$.

i	1	2	3	4	5	6	7	8	9	10
u_i	1	1	2	3	5	8	13	21	34	55
$w_i^{(0)}$				3/1	8/3	21/8	55/21...			
$w_i^{(1)}$				2/1	5/2	13/5	34/13...			

The secant method applied to $w_1^{(0)}$ and $w_2^{(0)}$ yields $21/8$ or $w_3^{(0)}$, applied to $w_1^{(0)}$ and $w_2^{(1)}$ yields $13/5$ or $w_3^{(1)}$ and applied to $w_1^{(1)}$ and $w_2^{(1)}$ yields $8/3$ or $w_2^{(0)}$ in agreement with equation (9). The Newton Raphson method applied to $w_1^{(0)}$ yields $8/3$ or $w_2^{(0)}$ and applied to $w_1^{(1)}$ yields $3/1$ or $w_1^{(0)}$ also as expected. Finally Aitken's transformation applied to $w_1^{(0)}$, $w_2^{(0)}$ and $w_3^{(0)}$ yields $55/21$ or $w_4^{(0)}$ and applied to $w_1^{(1)}$, $w_2^{(1)}$ and $w_3^{(1)}$ yields $21/8$ or $w_3^{(0)}$ in agreement with equation (11).

I am pleased to thank the referee for a useful comment.

REFERENCE

1. G. M. Phillips, Fibonacci numbers and Aitken sequences, *American Mathematical Monthly*, 91 (1984) 354–357.

A Cayley Theorem for Boolean Algebras

STEPHEN L. BLOOM[1]
Department of Computer Science, Stevens Institute of Technology, Hoboken, NJ 07030

ZOLTÁN ÉSIK
A. József University, Bolyai Institute, Szeged, Hungary

ERNEST G. MANES[2]
Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003

The well-known Cayley theorem for groups may be summarized as follows: any group is isomorphic to a group of transformations on some set. To make an

[1]Partially supported by NSF Grant CCR-8620250.
[2]Partially supported by NSF Grant CCR-8701272.

analogy below, we rephrase this theorem in two parts. Let X be a set and let $P(X)$ be the collection of all permutations of X—i.e. each g in $P(X)$ is a bijective function

$$g: X \to X.$$

We define two operations on $P(X)$: composition

$$(g \circ h)(x) := g(h(x));$$

and inverse

$$g^{-1}(x) := y \Leftrightarrow g(y) = x.$$

CAYLEY THEOREM, PART I. *If $P(X)$ is as above and if G is a subset of $P(X)$ which contains the identity function of X and is closed under composition and inverse then G is a group under these operations.*

Such groups are called transformation groups on X.

CAYLEY THEOREM, PART II. *If G is any group, there is some set X such that G is isomorphic to a transformation group on X.*

Now let X be a set and let f be a binary function on X. We say that the function f is *idempotent* if

$$f(x, x) = x$$

for all $x \in X$; f is *diagonal* (see [**4**]) if

$$f(f(x, y), f(u, v)) = f(x, v)$$

for all $x, y, u, v \in X$. If g is a binary function on X, we say that f and g *commute* if

$$f(g(x, y), g(u, v)) = g(f(x, u), f(y, v))$$

for all $x, y, u, v \in X$. We define some operations on the set $\mathrm{Bin}(X)$ of all binary functions on X:

$$\begin{aligned}(f \wedge g)(x, y) &:= f(g(x, y), y);\\ (f \vee g)(x, y) &:= f(x, g(x, y));\\ f'(x, y) &:= f(y, x);\\ \mathit{True}(x, y) &:= x;\\ \mathit{False}(x, y) &:= y.\end{aligned}$$

The following two-part theorem constitutes a "Cayley theorem for Boolean algebras."

PART I. (Cf. [**1**, §3]). *Let B be a subset of $\mathit{Bin}(X)$ with the following properties: B contains True and is closed under the operations of $\wedge$, $\vee$, and $'$. Then $\langle B, \wedge, \vee, ', \mathit{True}, \mathit{False}\rangle$ is a Boolean algebra providing each function in B is idempotent and diagonal and any two such functions commute.*

We call such Boolean algebras of binary functions on X "guard algebras on X" for reasons related to the concerns of [**3**, Chapter 4] and [**2**, §5]. (In computer science terminology, a "guard" is a Boolean condition which must hold before the procedure it guards can begin. Here we think of a guard P as represented by the binary function η on the set of functions computed by programs defined by $\eta(f, g) =$"if P then f else g." To understand our emphasis of the idempotency and diagonality properties, first represent the subset P of X by the partial function $p: X \to X$ defined by $x \mapsto x$ with domain P. If q is another such function corresponding to the subset Q, it is easily checked that P and Q are complementary if and only if $\eta(f, g) = f \circ p \vee g \circ q$ is idempotent and diagonal.)

PART II. *If $\langle B, \wedge, \vee, ', 1, 0\rangle$ is any Boolean algebra, there is a set X such that B is isomorphic to a guard algebra on X.*

The proof of Part I is routine. In order to prove part II, we introduce on any Boolean algebra B the binary functions if_b, for $b \in B$, defined by

$$if_b(x, y) := (b \wedge x) \vee (b' \wedge y).$$

We let the reader verify the following facts, which in ensemble constitute the proof of Part II.

$$\begin{aligned}
&if_b(x, x) = x;\\
&if_b(if_b(x, y), if_b(u, v)) = if_b(x, v);\\
&if_b(if_c(x, y), if_c(u, v)) = if_c(if_b(x, u), if_b(y, v));\\
&if_1(x, y) = True(x, y) = x;\\
&if_0(x, y) = False(x, y) = y;\\
&if_{b'}(x, y) = if_b(y, x) = (if_b)'(x, y);\\
&if_{b \wedge c}(x, y) = if_b(if_c(x, y), y) = (if_b \wedge if_c)(x, y);\\
&if_{b \vee c}(x, y) = if_b(x, if_c(x, y)) = (if_b \vee if_c)(x, y);\\
&if_b(1, 0) = b.
\end{aligned}$$

The last equation shows that the map $b \mapsto if_b$ is injective; the other equations show that each function if_b is idempotent and diagonal, that any two if_b, if_c commute and that the map is a Boolean homomorphism.

We thank George Bergman for providing a specific example of a 4-element Boolean algebra of binary functions which meets all the requirements for a guard algebra save that the nontrivial elements are not diagonal. This shows that the converse of Part I does not hold. If the definition of guard algebra were strengthened so that all its binary functions were required to be associative then both directions of Part I would hold and the corresponding version of Part II would also be true.

REFERENCES

1. G. Bergman, Actions of Boolean Rings on Sets, preprint, Univ. of Cal., Berkeley, CA 94720.
2. S. L. Bloom, Z. Ésik, Varieties of iteration theories, *SIAM J. Computing*, 17(1988) 939–66.
3. E. G. Manes, M. A. Arbib, Algebraic Approaches to Program Semantics, Springer-Verlag, 1986.
4. J. Plonka, Diagonal algebras, *Fundamenta Mathematicae*, LVIII (1966) 309–321.

Rouché's Theorem for Polynomials

MICHAEL FILASETA
Mathematics Department, University of South Carolina, Columbia, SC 29208

Rouché's Theorem (cf. Conway [**1**]) often permits one to determine something about the location of the roots or poles of a meromorphic function $f(z)$ by comparing $f(z)$ to another meromorphic function whose roots and poles are known. In the case of polynomials this is particularly useful to deduce the irreducibility of a given polynomial $f(z)$. Thus, for example, in this manner one can obtain Perron's Theorem [**4**]: if $f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$ has integer coefficients with $a_0 \neq 0$ and $|a_{n-1}| > 1 + |a_{n-2}| + \cdots + |a_0|$, then $f(z)$ is irreducible over the integers.

The purpose of this paper is to present a proof of Rouché's Theorem for polynomials. We shall take the simplest case (stated below) where the closed curve under consideration is the unit circle. It will be clear from the proof that more general curves could be used.

Before proceeding, we note that the proof given here makes use of the Fundamental Theorem of Algebra. On the other hand, one can prove the Fundamental Theorem of Algebra using Rouché's Theorem. Thus, one may view this paper as establishing that Rouché's Theorem for polynomials is equivalent to the Fundamental Theorem of Algebra (i.e., that each is easily derivable from the other). We add that proofs of the Fundamental Theorem of Algebra which do not make use of Rouché's Theorem can be found in Dickson [**2**], Grove [**3**], or Rudin [**5**], and that it is also an easy consequence of Liouville's Theorem.

ROUCHÉ'S THEOREM. *Let $f(z)$ and $g(z)$ be polynomials with complex coefficients and $C = \{z: |z| = 1\}$. If the strict inequality*

$$|f(z) + g(z)| < |f(z)| + |g(z)|$$

holds at each point on the circle C, then $f(z)$ and $g(z)$ must have the same total number of zeros (counting multiplicity) strictly inside C.

Proof. Let k be the number of roots of $f(z)$ inside C, and let t be the number of roots of $g(z)$ inside C. By the symmetry of the roles of $f(z)$ and $g(z)$ in the statement of the Lemma, it suffices to show that $k \leqslant t$. Assume $k > t$. We will obtain a contradiction by showing that there is a $\phi \in [0, 2\pi)$ such that

$$\left|f(e^{i\phi}) + g(e^{i\phi})\right| = \left|f(e^{i\phi})\right| + \left|g(e^{i\phi})\right|.$$

By the triangle inequality, it suffices to show that there is a $\phi \in [0, 2\pi)$ such that $\arg(f(e^{i\phi})) = \arg(g(e^{i\phi}))$, where for $z \in \mathbf{C}$ we take $\arg(z) \in (-\pi, \pi]$. For $\alpha \in \mathbf{C}$, $|\alpha| \neq 1$, define $w(\alpha; 0) = \arg(1 - \alpha)$ and for $\theta \in \mathbf{R} - \{0\}$ $w(\alpha; \theta) = \arg(e^{i\theta} - \alpha) + 2s\pi$ where $s = s(\theta)$ is an integer chosen so that $w(\alpha; \theta)$ is continuous on $\mathbf{R}$. Geometrically, $w(\alpha; \theta) - w(\alpha; 0)$ is a continuous representation of the angle with vertex at α and rays emanating through 1 and $e^{i\theta}$. Thus,

$$w(\alpha; 2\pi) - w(\alpha; 0) = \begin{cases} 2\pi & \text{if } \alpha \text{ is inside the circle } C \\ 0 & \text{otherwise.} \end{cases} \qquad (*)$$

Let a be the leading coefficient of $f(z)$ and b be the leading coefficient of $g(z)$. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the roots of $f(z)$ and $\beta_1, \beta_2, \ldots, \beta_m$ the roots of $g(z)$ (the

roots appearing as many times as their multiplicity). The inequality in the hypothesis of the theorem implies that each $|\alpha_j| \neq 1$ and each $|\beta_j| \neq 1$. Define

$$F(\theta) = \arg(a) + \sum_{j=1}^{n} w(\alpha_j; \theta) + 2u\pi,$$

where u is an integer chosen so that $F(0) \in (-2\pi, 0]$. Define

$$G(\theta) = \arg(b) + \sum_{j=1}^{m} w(\beta_j; \theta) + 2v\pi,$$

where v is an integer chosen so that $G(0) \in [F(0), F(0) + 2\pi)$. Thus, $F(\theta)$ and $G(\theta)$ are continuous on $\mathbf{R}$. Also, since

$$\arg(f(e^{i\theta})) = \arg\left(a\prod_{j=1}^{n}(e^{i\theta} - \alpha_j)\right) \equiv \arg(a) + \sum_{j=1}^{n} \arg(e^{i\theta} - \alpha_j) \pmod{2\pi},$$

we get that

$$F(\theta) \equiv \arg(f(e^{i\theta})) \pmod{2\pi}. \qquad (**)$$

Similarly,

$$G(\theta) \equiv \arg(g(e^{i\theta})) \pmod{2\pi}. \qquad (***)$$

Furthermore, by (*),

$$F(2\pi) = F(0) + 2k\pi \quad \text{and} \quad G(2\pi) = G(0) + 2t\pi.$$

Let $H(\theta) = F(\theta) - G(\theta)$. Then $H(0) = F(0) - G(0) \leqslant 0$ and $H(2\pi) = F(2\pi) - G(2\pi) = 2(k - t)\pi + F(0) - G(0) > 0$. By the Intermediate Value Theorem, there is a $\phi \in [0, 2\pi)$ for which $H(\phi) = 0$. By (**) and (***), we now get that $\arg(f(e^{i\phi})) = \arg(g(e^{i\phi}))$, completing the proof.

REFERENCES

1. J. B. Conway, Functions of One Complex Variable, second edition, Springer-Verlag, New York, 1978, p. 125.
2. L. E. Dickson, New First Course in the Theory of Equations, John Wiley & Sons, New York, 1939, p. 177.
3. L. C. Grove, Algebra, Academic Press, New York, 1983, p. 95.
4. O. Perron, Algebra II, de Gruyter, Leipzig, 1933, p. 35.
5. W. Rudin, Principles of Mathematical Analysis, third edition, McGraw-Hill, 1974, p. 184.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

A Simple Derivation of the Maclaurin Series for Sine and Cosine

DENG BO
Coal Exploration No. 174, Team of Gui Zhou, Zhi Jin, Gui Zhou Province, People's Republic of China

The standard derivations of the Maclaurin series for sine and cosine via differentiation are well-known. In this note I show how integration can be used to obtain the two series and prove that they converge to the sine and cosine functions.

Clearly, it suffices to do this for $x \geqslant 0$. Because $\cos x \leqslant 1$, $\int_0^x \cos y\,dy \leqslant \int_0^x 1\,dy$. That is, $\sin x \leqslant x$. From this it follows that $\int_0^x \sin y\,dy \leqslant \int_0^x y\,dy$. That is, $1 - \cos x \leqslant x^2/2!$. Continuing in this fashion, we obtain the inequalities

$$1 - \frac{x^2}{2!} \leqslant \cos x \leqslant 1$$

$$x - \frac{x^3}{3!} \leqslant \sin x \leqslant x$$

$$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} \leqslant \cos x \leqslant 1 - \frac{x^2}{2!} + \frac{x^4}{4!}$$

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \leqslant \sin x \leqslant x - \frac{x^3}{3!} + \frac{x^5}{5!}.$$

Now, still assuming $x \geqslant 0$, induction can be used to prove that $S(2r) \leqslant \cos x \leqslant S(2r-1)$ and $T(2r) \leqslant \sin x \leqslant T(2r-1)$ for $r = 1, 2, \ldots$, where

$$S(k) = 1 + \sum_{i=1}^{k-1} (-1)^i \frac{x^{2i}}{(2i)!} \quad \text{and} \quad T(k) = \sum_{i=0}^{k-1} (-1)^i \frac{x^{2i+1}}{(2i+1)!}.$$

The upper and lower bounds differ by $x^m/m!$, which, for fixed x, can easily be shown to approach 0 as $m \to \infty$.

An Infinite Product Expansion for the Square Root Function

ERIC WINGLER
Department of Mathematical and Computer Sciences, Youngstown State University, Youngstown, OH 44555

Occasionally when we try to solve a problem by one method we discover another method that works better than the one we had in mind. Consider the

problem of calculating the square root of a number using the Maclaurin series for $\sqrt{1+x}$,

$$\frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \cdots. \tag{1}$$

Since this series has a radius of convergence equal to 1 (see [**2**, §8.6]), we can only use it directly to find square roots of numbers in the interval $[0, 2]$. However, it is readily seen that to calculate $\sqrt{10}$, say, all that is needed is to compute $\sqrt{10/9}$ using (1) and then to multiply by 3. But even if we are using this series to compute square roots of numbers in the interval $[0, 2]$, the process can be quite lengthy if we desire a reasonable degree of accuracy. For example, it takes approximately 10,000 terms to approximate $\sqrt{2}$ to within 3×10^{-7}.

Is there any way to improve convergence? It is in answering this question that we discover something remarkable. Obviously the series (1) converges best for values of x close to 0. So in calculating $\sqrt{2}$ we try to express 2 as a product of a square and a number closer to 1 than 2 is. For example,

$$\sqrt{2} = \frac{4}{3}\sqrt{2 \cdot \left(\frac{3}{4}\right)^2} = \frac{4}{3}\sqrt{1 + \frac{1}{8}}.$$

This works so well that we are tempted to try this again:

$$\sqrt{2} = \frac{4}{3}\sqrt{9/8} = \frac{4}{3} \cdot \frac{18}{17}\sqrt{\frac{9}{8}\left(\frac{17}{18}\right)^2} = \frac{24}{17}\sqrt{1 + \frac{1}{288}}.$$

But where do we stop? We don't! This method works so well that there is no need to return to using power series.

The trick is to find the appropriate thing to factor out. One way of choosing things to factor out is suggested by the following:

$$\sqrt{1+x} = \frac{2x+2}{x+2}\sqrt{(1+x)\left(\frac{x+2}{2x+2}\right)^2} = \frac{2x+2}{x+2}\sqrt{1 + \frac{x^2}{4x+4}}. \tag{2}$$

Thus if we let $a_0 = x$, $a_{n+1} = a_n^2/(4a_n + 4)$, and $b_n = (2a_n + 2)/(a_n + 2)$ for each integer $n \geqslant 0$, then we have

$$\sqrt{1+x} = \prod_{n=0}^{\infty} b_n.$$

Since the numbers a_n converge quadratically to 0, this product converges rapidly. For example, if we use $x = 1$, we have $|\sqrt{2} - b_0 b_1 b_2 b_3| < 4 \times 10^{-12}$.

As you might have guessed from (2), the way to improve convergence is to find two functions f and g such that

$$\sqrt{1+x} = f(x)\sqrt{1+g(x)}, \tag{3}$$

where g is of smaller order than x as x approaches 0. We can find good choices for f and g by using (1). Of course, once f is chosen, then g is automatically determined by (3). Specifically,

$$g(x) = \frac{1+x}{f(x)^2} - 1. \tag{4}$$

First, observe that if $f(x) - \sqrt{1 + x}$ is $O(x^n)$ as x approaches 0, then $1 + x = f(x)^2 + O(x^n)$, so that $g(x) = O(x^n)$ by (4). So if the first n terms of the Maclaurin series for f match those of $\sqrt{1 + x}$, then $g(x) = O(x^n)$.

The nicest choices for f seem to be rational functions whose numerator and denominator are of the same degree. We list here some specific cases.

$$f(x) = \frac{3x + 4}{x + 4}; \qquad g(x) = \frac{x^3}{(3x + 4)^2} \tag{5}$$

$$f(x) = \frac{5x^2 + 20x + 16}{x^2 + 12x + 16}; \qquad g(x) = \frac{x^5}{(5x^2 + 20x + 16)^2} \tag{6}$$

$$f(x) = \frac{7x^3 + 56x^2 + 112x + 64}{x^3 + 24x^2 + 80x + 64}; \quad g(x) = \frac{x^7}{(7x^3 + 56x^2 + 112x + 64)^2}. \tag{7}$$

These functions are easy to obtain. Simply find the coefficients for two polynomials $P(x)$ and $Q(x)$ that will make the first n terms of the Maclaurin series for $P(x) - Q(x)(\sqrt{1 + x})$ disappear. In general if deg P = deg $Q = k$, then $n = 2k + 1$. Then $f(x) = P(x)/Q(x)$ and $g(x) = mx^{2k+1}/P(x)^2$, where $m = 2Q(1)^2 - P(1)^2$. The functions f obtained by using this procedure are called the Padé approximants of $\sqrt{1 + x}$. The functions in (5), (6), and (7) are the Padé approximants denoted by $f_{1,1}$, $f_{2,2}$, and $f_{3,3}$ in [**3**, Chap. XX] (also see [**1**]).

For any functions f and g satisfying (3) and such that $g(x) = O(x^n)$ with $n \geqslant 2$, the method for calculating $\sqrt{1 + x}$ is the same. Let $a_0 = x$, $a_{n+1} = g(a_n)$, and $b_n = f(a_n)$ for each integer $n \geqslant 0$. Then

$$\sqrt{1 + x} = \left(\prod_{n=0}^{k} b_n\right)\sqrt{1 + a_{k+1}}.$$

So for $x \geqslant 0$

$$\left|\sqrt{1 + x} - \prod_{n=0}^{k} b_n\right| = \left(\prod_{n=0}^{k} b_n\right)\left|\sqrt{1 + a_{k+1}} - 1\right| \leqslant \frac{a_{k+1}}{2}\sqrt{1 + x},$$

since

$$\prod_{n=0}^{k} b_n \leqslant \sqrt{1 + x} \text{ and } \sqrt{1 + a_{k+1}} - 1 \leqslant \frac{a_{k+1}}{2}.$$

It should be pointed out that all of the functions g in (5), (6), and (7) are contractive; that is, $|g(x)| < \alpha|x|$ for some $\alpha < 1$ and all $x \geqslant 0$. For example, using the function in (5), we have for $x \geqslant 0$

$$|g(x)| = \frac{|x|^3}{|3x + 4|^2} \leqslant \frac{|x|^3}{9x^2} = \frac{|x|}{9}.$$

Therefore, the sequence (a_n) will converge rapidly to 0. In fact, if (7) is used with $x = 1$, then we get

$$|\sqrt{2} - b_0 b_1 b_2| < 5 \times 10^{-255}.$$

Finally, we remark that this procedure can be used for other roots as well; for example,

$$\sqrt[3]{1+x} = \left(\frac{2x+3}{x+3}\right)\sqrt[3]{1+\frac{2x^3+x^4}{(2x+3)^3}}\,.$$

REFERENCES

1. G. A. Baker and P. Graves-Morris, Padé approximants, Encyclopedia of Mathematics and its Applications, Vols. 13 and 14, Addison-Wesley, Reading, Mass., 1981.
2. Richard R. Goldberg, Methods of Real Analysis, second ed., Wiley, New York, 1976.
3. H. S. Wall, Continued Fractions, Chelsea, New York, 1973.

Some Models of Plane Geometries*

BRANKO GRÜNBAUM
Department of Mathematics, University of Washington, Seattle, WA 98195

JAN MYCIELSKI
Department of Mathematics, University of Colorado, Boulder, CO 80309

Undergraduate courses on the foundations of geometry need examples different from the traditional Euclidean plane in order to give students experience in dealing with axiomatic systems. In most texts, such examples are given in the discussion of the axioms of incidence; they consist, almost invariably, of finite geometries or related systems (see, e.g., [**3**]). In our experience, such examples have very little value for most of the mathematically rather unsophisticated students that take the geometry-for-teachers courses. Moreover, as soon as one reaches the axioms of order (in any of their embodiments), these examples have to be abandoned and there are very few others to take their place. (The text [**6**] by Millman and Parker is a notable exception to the tendency of presenting only a very small collection of models. Gans [**2**] also describes a model rarely mentioned by other writers; it is the topic of a recent paper [**5**] in the MONTHLY. For other discussions of this and related models see [**4**, p. 88] and the references given there.) The aim of the present note is to provide several models the first author has used in such courses. These models seem to lead to better understanding of axiomatics by the students—probably because they are, in a psychological sense, more like real geometry than the usual examples. (This experience seems to be in agreement with the findings of Nisbett *et al.* [**7**].) It is hoped that the models may provide some amusement and some surprises for the teacher as well. The exploration of the models can be particularly satisfying using a Macintosh computer and such software as MacDraw.

All the models are set in the usual Euclidean plane $\mathbb{E}^2$, the "ordinary plane" of the student's experience; the points of each of our geometries coincide with the

*Research supported in part by National Science Foundation grant DMS-8620181.

points of $\mathbb{E}^2$. If a more formal setting is desired then, just as with the Euclidean geometry, the points P can be taken as ordered pairs $P = (x, y)$ of real numbers. The geometries we discuss differ from each other by their sets of lines.

Geometry **G1**. In this geometry a "line" is any one of the following objects (which visually have quite different shapes; see FIGURE 1). Let U be a fixed (Euclidean) parabola, the axis of which we shall assume to be vertical; in the analytic formulation, we can take U as the set given by $U = \{(x, y): y = x^2\}$. Then the lines of **G1** are:

(i) all the translates of U;

(ii) all the vertical Euclidean lines (given by $\{(x, y): x = c\}$).

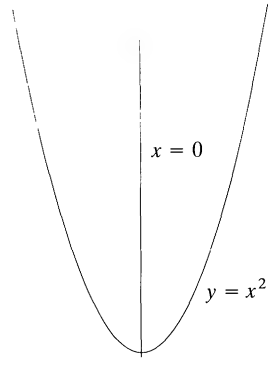


FIG. 1. Examples of lines in the geometry **G1**.

Geometry **G2**. Here a "line" is any of the following objects (see FIGURE 2):

(i) a vertical Euclidean line;

(ii) a horizontal Euclidean line;

(iii) a translate of the first-quadrant branch L of a fixed equilateral hyperbola with vertical and horizontal asymptotes; analytically, this can be taken as

$$L = \{(x, y): x > 0, y > 0, xy = 1\};$$

(iv) a translate of the reflection L^* of L in the y-axis, that is,

$$L^* = \{(x, y): x < 0, y > 0, xy = -1\}.$$

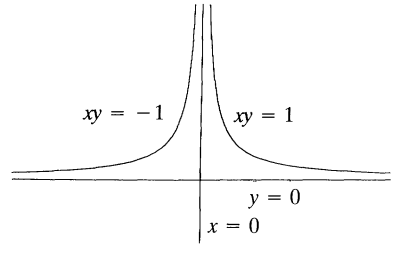


FIG. 2. Examples of lines in the geometry **G2**.

Geometry **G3**. Here a "line" is any of the following objects (see FIGURE 3):

(i) a horizontal Euclidean line;

(ii) a translate of the branch L to the left of the Y-axis of a hyperbola with asymptotes at $\pm 60°$ to the x-axis; analytically, this can be taken as

$$L = \{(x, y): x < 0, 3x^2 - y^2 = 4\};$$

(iii) any object obtained by rotating one of those in (i) or (ii) through 120° or 240°.

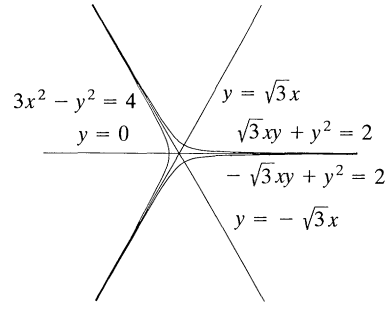


FIG. 3. Examples of lines in the geometry **G3**.

To study these geometries, two ways are open; it is best to use them both. On the one hand, practical experience concerning the meaning of the axioms, and whether they are satisfied or not, can be obtained by experimentation in drawing. "Rulers" for the "lines" are quite easily prepared; if a photocopying machine that accepts cardboard is available, the teacher can provide such "rulers" (ready to be cut out) to the students. Otherwise copies on paper should be pasted by students onto cardboard, and then cut. To assure that only translates of a fixed parabola or hyperbola are used, it is convenient to draw on ruled (or engineering) paper, and make a slit in the template, which enables one to check alignment. On the other hand, one can use the equations to proceed in an algebraic way to verify the various claims, and to settle disputes concerning the drawing (which inevitably arise due to the limited precision of the practical procedures). Clearly, each "parabolic" line L in **G1** will have an equation of the type $y = (x - a)^2 + b$, where (a, b) is the apex of L. Similarly, the "hyperbolic" lines of **G2** will have equations of the type $(x - a)(y - b) = \pm 1$, with suitable inequalities for x and y; the appropriate relations for **G3** are also easily found.

Naturally, for a model of a geometry one needs not only points and lines, but incidences, order (or betweenness), and so on. In class these can be either imposed by the teacher or "discovered" by the students. The advantage of the models under discussion over the usual ones is that even the study of axioms of incidence and order provides meaningful experiences, as do questions on parallelism and related topics. Experimenting first, and verifying later (by computation or in some other way), will show that the three models satisfy all the usual axioms on incidence and order. Also, by defining the distance between two points as being equal to the distance in the Euclidean sense along the "segment" determined by the points, it

is easily seen that the congruence and continuity axioms are also satisfied, with the exception of those that deal with triangles. As far as parallels are concerned, **G1** behaves like the Euclidean plane (in that for each line and each point there is a unique parallel to the line through that point) while **G2** and **G3** do not. An interesting difference between **G2** and **G3** is that in the former there are lines which for some points have unique parallels; this does not happen in **G3**, which is therefore a "hyperbolic" model. From an axiomatic point of view, model **G2** is interesting since in it, in contrast to the situation usually considered in axiomatic treatments, the failure of the Playfair axiom for one line and one point does not imply its failure for all such pairs; thus **G2** shows that the theorem about the uniform behavior of line-point pairs does not have to hold if one of the congruence axioms is not assumed.

With that much established, one can investigate such topics as the numbers of regions, segments, and intersection points of finite families of lines every two of which intersect, but no three have a common point, or of more general families. Or one can investigate whether such results as the theorems of Pappus or Desargues (or suitable special cases, which may be easier to formulate and to check) are valid; see FIGURE 4.

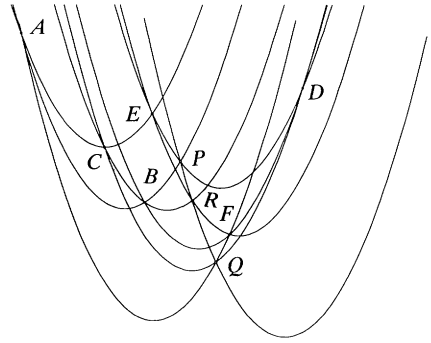


FIG. 4. Illustration of the theorem of Pappus in the geometry **G1**. The vertices of the hexagon *ABCDEF* alternate between two lines, hence the intersection points P, Q, R of pairs of opposite sides of the hexagon are collinear.

An ambitious teacher may define the midpoint of a segment in **G1** (as the usual midpoint for vertical segments, and as the point of the parabola at which the tangent is parallel to the chord determined by the endpoints of the segment for segments that are parabolic arcs) and prove (or have students prove) the concurrence of the medians of a triangle (see FIGURE 5), and other simple facts. An even more ambitious project is to verify that in **G1** a distance function (different from the one mentioned above) can be defined between pairs of points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ by the (somewhat intimidating) formula

$$d(P_1, P_2) = \left((x_1 - x_2)^2 + \left(y_1 - y_2 - x_1^2 + x_2^2\right)^2\right)^{1/2}.$$

Then it is possible to establish that with this distance function the geometry **G1**

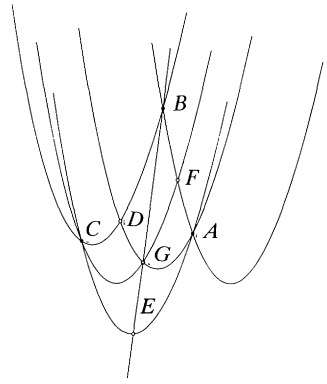


FIG. 5. In geometry **G1**, the medians AD, BE, and CF of a triangle ABC meet at the centroid G.

satisfies *all* the axioms of congruence and of continuity—in short, that **G1** is a model of the Euclidean geometry!

A simpler way to see that **G1** is just an unusual version of the usual Euclidean plane is to consider the mapping ϕ of the plane of **G1** to the Euclidean plane $\mathbb{E}^2$ given by $\phi(x, y) = (x, y - x^2)$. It is clear that ϕ is a continuous bijection, and that it maps the lines of **G1** onto lines of the Euclidean plane, and vice versa. Thus **G1**, as we have defined it, is just a line-preserving homeomorphic image of the Euclidean plane—but this does not have to be told to the students until an appropriate later time.

Since in **G2** some point-line pairs behave as in the Euclidean plane and some as in the hyperbolic plane, this geometry is not a homeomorphic line-preserving image of any of the standard geometries. As mentioned earlier, the geometry **G3** behaves as expected from the hyperbolic plane, and it is natural to ask whether it can be metrized in such a way as to be homeomorphic to the usual hyperbolic plane (under a line-preserving mapping). There seem to be no direct methods available at present to decide such metrization problems, or of finding a metric if it exists. However, an indirect proof for the essential difference between **G3** and the hyperbolic plane (and thus the nonexistence of a line-preserving homeomorphism between them) follows from the observation made by Walter Taylor that in **G3** some angles are not "angles of parallelism." Indeed, there is no common parallel to the positive rays of $y = \sqrt{3}x$ and $y = -\sqrt{3}x$, in contrast to the situation regarding any two non-opposite rays in the hyperbolic plane.

As an additional project one could consider geometries analogous to those described above, but with other suitable curves substituted for the parabolas or hyperbolas. As examples one could use $y = \cosh x$ instead of the parabola in **G1**, or translates of $y = e^x$ and $y = e^{-x}$ instead of the branches of the hyperbolas in **G2**. In fact, only certain mild monotonicity conditions have to be imposed on the curves used to obtain geometries satisfying the axioms of incidence, order and continuity, as well as most of the axioms of congruence. Moreover, there is no need for the curves in different sectors to be congruent; they can be chosen with a large degree of freedom. Another way of extending the family of models is to take n equi-inclined (Euclidean) straight lines through the origin (with $n \geqslant 2$), and a hyperbola H (or other suitable curve) having as asymptotes two rays (determined

by these lines) which enclose an angle of $(n-1)\pi/n$. The lines of the geometry **Gn** are translates of the n lines we started with, as well as translates of H and of the hyperbolas obtained from it by certain rotations. For odd n, it is most convenient to choose the n rotations of H through multiples of $2\pi/n$. For even n, the rotations through $k\pi/n$ with $0 \leqslant k \leqslant n-1$ can be taken. (For $n = 2$ or 3 these definitions agree with the previous ones.) Other sets of rotations can also be used, but tend to lead to messy case distinctions.

In contrast to this, we shall prove that **G1** is in a certain sense unique. To formulate our result precisely, let $\mathscr{L}$ denote the family of all straight lines in the Cartesian plane $\mathbb{R}^2$, and let $\mathscr{M}$ denote the family consisting of all vertical lines in $\mathbb{R}^2$ and all curves that are translation-congruent to one given by the equation $y = p(x)$, where p is a function defined for all $x \in \mathbb{R}$. We say that θ is an isomorphism from $\mathscr{L}$ to $\mathscr{M}$ if θ is a bijection of $\mathbb{R}^2$ onto itself which induces a bijection of $\mathscr{L}$ onto $\mathscr{M}$.

THEOREM. *If there exists a continuous isomorphism θ from $\mathscr{L}$ to $\mathscr{M}$ and $p(x)$ is differentiable for all x, then p is a polynomial of degree two.*

We do not know whether the assumptions of differentiability of p and continuity of θ are essential.* (We know that measurability or the property of Baire of θ would suffice.) The following is a well-known result on collineations (see, e.g., [**1**, p. 156] for a proof of the projective geometry formulation; the result is often called the "Fundamental Theorem of Projective Geometry"). It suggests that such assumptions about p or θ could be superfluous.

LEMMA. *Every bijection φ of $\mathbb{R}^2$ onto itself which induces a bijection of $\mathscr{L}$ onto itself is affine linear, that is,*

$$\varphi(x, y) = (ax + by + u, cx + dy + v),$$

where $ad - bc \neq 0$.

PROOF (in outline). Let $\psi = \alpha \circ \varphi$, where α is affine linear. We can choose α such that ψ fixes three non-collinear points p_0, p_1, p_2. Interpreting the straight lines $\overrightarrow{p_0p_i}$ $(i = 1, 2)$ in the natural way as copies of the field of real numbers, with $p_0 = 0$ and $p_i = 1$, we can verify by means of familiar constructions that the restriction of ψ to $\overrightarrow{p_0p_i}$ is an automorphism of this field. Since the field of real numbers has no automorphisms different from the identity, it follows that $\psi(p) = p$ for all $p \in \overrightarrow{p_0p_i}$. Then it is easy to check that $\psi(p) = p$ for all $p \in \mathbb{R}^2$. Hence $\varphi = \alpha^{-1}$ which proves the Lemma. □

Proof of the Theorem. Let $\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a continuous isomorphism from $\mathscr{L}$ to $\mathscr{M}$. Then for any affine linear automorphism α of $\mathscr{L}$, the composition $\kappa = \theta \circ \alpha$ is still a continuous isomorphism from $\mathscr{L}$ to $\mathscr{M}$. Since the family of vertical straight lines is a maximal family of parallel lines both in $\mathscr{L}$ and in $\mathscr{M}$, we can choose α such that κ maps this family onto itself. Hence κ is of the form $\kappa(x, y) = (f(x), g(x, y))$, where f and g are continuous. Refining our choice of α we can assume that $f(0) = 0$.

*V. Faber and J. Mycielski have shown that the continuity of p and θ suffices [to appear].

Now consider the one-parameter group of automorphisms of $\mathscr{M}$ given by $\tau_t(x, y) = (x + t, y)$. Then the maps $\tau_t' = \kappa^{-1} \circ \tau_t \circ \kappa$ constitute a one-parameter group of automorphisms of $\mathscr{L}$. By the lemma, all τ_t' are affine linear, and moreover each τ_t' preserves the family of vertical lines but does not fix any of them for $t \neq 0$. Hence

$$\tau_t'(x, y) = (x + h(t), a(t)x + b(t)y + c(t)), \qquad (*)$$

where h, a, b and c are some functions and $h(s + t) = h(s) + h(t)$ for all $s, t, \in \mathbb{R}$. Since $\kappa \circ \tau_t' = \tau_t \circ \kappa$, we have $f(x + h(t)) = f(x) + t$. From $f(0) = 0$, putting $x = 0$, we get $f \circ h(t) = t$ for all t. Hence $h = f^{-1}$ and h is continuous. By the theorem of Cauchy about the linearity of continuous additive functions it follows that $h(t) = at$ for some $a \neq 0$. (At this point it would have been sufficient to know that h is measurable, or has the property of Baire.) Hence $f(x) = x/a$. Refining again our choice of α we can assume that $a = 1$. So we get $\kappa(x, y) = (x, g(x, y))$.

Now consider the one-parameter group of automorphisms of $\mathscr{L}$ and of $\mathscr{M}$ given by $\sigma_t(x, y) = (x, y + t)$. This is the only one-parameter group of automorphisms of $\mathscr{L}$ (and of $\mathscr{M}$) which preserve every vertical line and acts without fixed points for $t \neq 0$. Hence the map $\sigma_t \mapsto \sigma_t' = \kappa^{-1} \circ \sigma_t \circ \kappa$ must be an automorphism of this group. Therefore $\sigma_t'(x, y) = (x, y + k(t))$, where k is continuous and additive. So again $k(t) = at$, where $a \neq 0$. Since $\kappa \circ \sigma_t' = \sigma_t \circ \kappa$, we have $g(x, y + k(t)) = g(x, y) + t$. Hence $g(x, y) = g(x, 0) + y/a$. Refining again our choice of α we can assume that $a = 1$ and then, by the definition of $\mathscr{M}$, the graph of $g(x, 0)$ must be translation-congruent to the graph of $p(x)$. It follows that $g(x, y) = q(x) + y$, where $q(x) = c_0 + p(x + c_1)$ and c_0 and c_1 are suitable constants.

Hence we have $\kappa(x, y) = (x, y + q(x))$ and, returning to the transformations τ_t' defined above, we get

$$\begin{aligned}
\tau_t'(x, y) &= \kappa^{-1} \circ \tau_t \circ \kappa(x, y) \\
&= \kappa^{-1} \circ \tau_t(x, y + q(x)) \\
&= \kappa^{-1}(x + t, y + q(x)) \\
&= (x + t, y + q(x) - q(x + t)).
\end{aligned}$$

Comparing with $(*)$ we get $b(t) = 1$ and

$$q(x) - q(x + t) = a(t)x + c(t).$$

Then putting $x = 0$ we get $q(0) - q(t) = c(t)$. Hence, for $x = 1$,

$$a(t) = q(1) - q(1 + t) - q(0) + q(t),$$

and

$$q(x) - q(x + t) = (q(1) - q(1 + t) - q(0) + q(t))x + q(0) - q(t).$$

Differentiating relative to t we get

$$-\dot{q}(x + t) = (\dot{q}(t) - \dot{q}(1 + t))x - \dot{q}(t),$$

and, for $t = 0$,

$$\dot{q}(x) = (\dot{q}(1) - \dot{q}(0))x + \dot{q}(0).$$

Integrating in x we see that $q(x)$ is a polynomial of degree at most 2, and the

same is true for p. Since a linear function p does not satisfy the assumptions of the Theorem, it follows that p is a polynomial of degree 2. □

REFERENCES

1. R. Artzy, Linear Geometry, Addison-Wesley, Reading, Mass., 1965.
2. D. Gans, Transformations and Geometries, Appleton-Century-Crofts, New York, 1968.
3. M. J. Greenberg, Euclidean and Non-Euclidean Geometries, second ed., Freeman, San Francisco, 1980.
4. B. Grünbaum, Arrangements and Spreads. CBMS Regional Conference Series in Mathematics, Number 10, Amer. Math. Soc., Providence, R.I., 1972.
5. A. Mader, A Euclidean model for Euclidean geometry, this MONTHLY, 96 (1989), 43–49.
6. R. S. Millman and G. D. Parker, Geometry: A Metric Approach With Models, Springer-Verlag, New York, 1981.
7. R. E. Nisbett, G. T. Fong, D. R. Lehman, and P. W. Cheng, Teaching reasoning, *Science*, 238 (1987) 625–631.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, JOHN R. ISBELL, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, FREDERICK W. LUTTMAN, MARVIN MARCUS, FRANK B. MILES, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

An asterisk () after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before April 30, 1991 at the MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3403. *Proposed by Paul Erdös, Hungarian Academy of Sciences, Budapest.*

It is well known that the maximum size of a subset of $\{1, 2, \ldots, n\}$ containing no pair i, j with $i|j$ is $\lfloor (n+1)/2 \rfloor$. Prove that the maximum size of a subset of $\{1, 2, \ldots, n\}$ containing no pair i, j with $i|2j$ is $4n/9 + O(\log n)$ for large n.

E 3404. *Proposed by the editors.* (*A modification of a problem proposed by the late Reuven Gurevič.*)

Suppose f is a function from R to R such that $\sum f(a_n)$ is convergent whenever $\sum a_n$ is a convergent series of real terms. Prove that f is differentiable at the origin.

E 3405. *Proposed by Charles Vanden Eynden, Illinois State University, Normal, IL.*

Suppose m, n are integers with $1 < m \leqslant n$. Let S_n be the thick-walled box obtained by removing the cube

$$\{(x, y, z): 1 < x, y, z < n - 1\}$$

from the cube

$$\{(x, y, z): 0 \leqslant x, y, z \leqslant n\}.$$

For which pairs of integers m, n can S_n be constructed with $m \times 1 \times 1$ blocks?

E 3406. *Proposed by Jeffrey Shallit, Dartmouth College, Hanover, NH.*

Consider three circles in the plane that intersect to form seven bounded regions. In each region there is a token that is white on one side and black on the other. At any stage the following two operations are permissible: (a) we can invert (flip over) all four tokens inside one of the three circles, or (b) we can invert those tokens showing black inside one of the three circles so that afterwards all tokens in that circle show white. From the starting configuration in which all tokens show white, can we reach the configuration in which all tokens show white except that the central region common to the three discs shows black?

E 3407. *Proposed by Clark Kimberling, University of Evansville, Evansville, IN.*

Suppose ABC is a given triangle. Prove the existence of a triangle that is in perspective with every antipedal triangle of ABC. (If P is a point not collinear with any two of the the points A, B, C, the lines through A, B, C perpendicular to PA, PB, PC respectively form a triangle called the *antipedal triangle of* P with respect to ABC. Two triangles $A_1B_1C_1$ and $A_2B_2C_2$ are said to be *in perspective* if the lines A_1A_2, B_1B_2, C_1C_2 are concurrent.)

E 3408. *Proposed by Juan V. Savall and Jesús Ferrer, Oliva, Valencia, Spain*

For each positive integer k let $f(k)$ denote the number of triangles with sides of integer length whose area is k times the perimeter. It is well-known (cf. E2420 [1973, 691; 1974, 662]) that $f(1) = 5$. Obtain an upper bound for $f(k)$ in terms of k.

The analogous problem for *right* triangles appeared as Problem 1447 in Crux Mathematicorum, 15 (1989) 148.

SOLUTIONS OF ELEMENTARY PROBLEMS

Conncected Graphs with Every Edge in a Triangle

E 3255 [1988, 259]. *Proposed by Paul Erdös, Hungarian Academy of Sciences.*

What is the minimum number of edges in a connected n-vertex graph such that every edge belongs to a triangle?

Solution by David Callan, University of Bridgeport, Bridgeport, CT. Say that G is a T-graph if it is connected and each edge belongs to a triangle. Let $G = (V, E)$ be an n-vertex T-graph with the minimum number a_n of edges, and let $b_n = \lfloor (3n - 2)/2 \rfloor = \lceil 3(n - 1)/2 \rceil$. Note that $a_3 = b_3 = 3$. If n is odd, then $(n - 1)/2$ triangles sharing a common vertex form a T-graph with b_n edges. If n is even, then a similar structure with $(n - 2)/2$ triangles together with an additional vertex adjacent to the endpoints of one edge of this forms a T-graph with b_n edges. Hence $a_n \leqslant b_n$.

To see the opposite inequality, consider an n-vertex T-graph G with a_n edges. Since $a_n < 3n/2$, the sum of degrees in G is less than $3n$, so G must have a vertex

v_1 of degree 2. Let v_2, v_3 be the neighbors of v_1. Since each edge belongs to a triangle, $v_2v_3 \in E(G)$. If $n > 3$, form a smaller T-graph $G' = (V', E')$ as follows. Delete v_1 from G, and if v_2, v_3 have no common neighbor in G other than v_1, then contract v_2v_3. Every edge of G' belongs to a triangle that contained it in G. The change reduces the number of vertices by 1 or 2 and reduces the number of edges by at least $3/2$ times the reduction in the number of vertices. Repeating this process leads eventually to a 3-vertex graph, which must be a triangle. It follows that $a_n \geq 3 + 3(n - 3)/2 = 3(n - 1)/2$. Since a_n is an integer, we conclude $a_n \geq b_n$.

Editorial comment. F. Galvin points out that $\lfloor (3n - 2)/2 \rfloor$ is also the minimum number of edges in an n-vertex graph satisfying the stronger condition that every pair of distinct vertices has a common neighbor (see Ex. 2.4 of Béla Bollobás, *Combinatorics*, Cambridge University Press, 1986, p. 9). He and J. T. Ward (independently) also generalized the problem, replacing the word "triangle" in the problem statement by the word "m-gon". More generally, we can consider the question with "triangle" replaced by an arbitrary graph T with m edges and r vertices and ask whether the formula $m(n - 1)/(r - 1)$ is the correct answer when $n \equiv 1 \bmod r - 1$. This is not true in general if T has a vertex of degree less than $m/(r - 1)$. In this case one does better by appending $n - r$ vertices having the minimum degree to a single copy of T. However, if T is a regular graph, then we ask whether edge-disjoint copies of T sharing a single vertex always provide an optimal construction. S. F. Barger proved that this is true when T is the complete graph on r vertices, and more generally gave the complete solution for all congruence classes of $n \bmod r - 1$ in that case.

Solved also by S. F. Barger, P. A. Catlin & A. M. Hobbs, C. Coleman, J. Davis, W. Fenton, K. Fogarty, F. Galvin, D. C. Kay, O. P. Lossers (The Netherlands), D. P. Mehendale & M. R. Modak (India), L. Nykáky (France), A. Palmer (student), A. Pedersen (Denmark), R. G. Powers & K. D. Wallace, R. B. Richter, A. Riese, E. Schmeichel, M. Shimure, J. H. Steelman, J. T. Ward, K. Yamaoka & A. Riskin, and the proposer. Several incorrect solutions were received.

The Three Centroids of a Triangle

E 3283 [1988, 762]. *Proposed by Orrin Frink, Pennsylvania State University, State College, PA.*

Every simple closed polygon in the plane has three centroids, namely the centroid of its vertex set, the centroid of its boundary, and the centroid of its interior. In general all three are distinct.

(a) In the case of a triangle, show that these three centroids coincide if and only if the triangle is equilateral.

(b) Which of the three centroids are affine invariants?

Solution by Allan Pedersen, Søborg, Denmark. (a) It is known and easy to prove that the interior centroid C_i of any triangle coincides with its vertex centroid C_v, and that this is the intersection of the medians. The boundary centroid C_b of a triangle ABC is the barycenter of mass distributed along its sides with constant, say unit, density. It is a folk theorem that, given a mass distribution and a finite partition of this, the barycenter is the same as that obtained after having concentrated the masses of the parts at the various barycenters. Hence C_b is the

barycenter of a mass a at the midpoint A' of BC, a mass b at the midpoint B' of AC, and a mass c at the midpoint C' of AB. C_b is on AA' if and only if the barycenter of the masses at B' and C' is at the midpoint of $B'C'$, i.e., precisely when $b = c$. Thus C_b is on all the medians if and only if $a = b = c$.

(b) Let $\mathbf{P}$ be a simple closed polygon. Denote by $\mu_v^{\mathbf{P}}, \mu_b^{\mathbf{P}}, \mu_i^{\mathbf{P}}$ the uniform distributions of a unit mass on the vertices, the boundary, and the interior, respectively. The corresponding barycenters are the centroids $C_v^{\mathbf{P}}, C_b^{\mathbf{P}}, C_i^{\mathbf{P}}$. For $x \in \{v, b, i\}$, $C_x^{\mathbf{P}}$ is an affine invariant if and only if $f(C_x^{\mathbf{P}}) = C_x^{f(\mathbf{P})}$ for any $\mathbf{P}$ and any affine transformation f. By linearity of f, $f(C_x^{\mathbf{P}})$ is the barycenter of the mass distribution $f(\mu_x^{\mathbf{P}})$ on $f(\mathbf{P})$ transferred from $\mathbf{P}$ by means of f. Clearly $f(\mu_v^{\mathbf{P}}) = \mu_v^{f(\mathbf{P})}$. Also, since the Jacobian of f is constant, $f(\mu_i^{\mathbf{P}}) = \mu_i^{f(\mathbf{P})}$. Accordingly, $C_v^{\mathbf{P}}$ and $C_i^{\mathbf{P}}$ are affine invariants.

However, $f(\mu_b^{\mathbf{P}})$ is supported on the sides of $f(\mathbf{P})$, uniformly distributed on each side, but in general with different densities, since ratios of sides are not affine invariants. Therefore, it is not to be expected that $C_b^{\mathbf{P}}$ is an affine invariant. For an explicit example, choose $\mathbf{P}$ to be an equilateral triangle, and choose f so that $f(\mathbf{P})$ is not equilateral. Then $C_b^{\mathbf{P}} = C_v^{\mathbf{P}}$ and $f(C_b^{\mathbf{P}}) = C_v^{f(\mathbf{P})}$, but (a) implies $C_b^{f(\mathbf{P})} \neq C_v^{f(\mathbf{P})}$.

Editorial comment. As shown by O. P. Lossers, the centroid of the boundary of a triangle is the center of the inscribed circle of the triangle with vertices at the midpoints of the sides. A proof is given in R. A. Johnson, *Modern Geometry*, (Houghton-Mifflin Co., 1929), Theorem 412. Johnson calls the inscribed circle of the midpoint triangle the "Spieker circle".

Solved also by B. Kirchheim (Czechoslovakia), O. P. Lossers (The Netherlands), J.-M. Monier (France), and A. Nijenhuis. One incorrect solution was received.

Divisibility of Multinomial Coefficients

E 3313 [1989, 253]. *Proposed by Amir Daneshgar (student), Sharif University of Technology, Teheran, Iran.*

Prove that the number of terms in the expansion of $(x_1 + x_2 + \cdots + x_m)^n$ whose coefficients are not divisible by a given prime number p is

$$\prod_{i=0}^{k} \binom{a_i + m - 1}{m - 1},$$

where

$$n = \sum_{j=0}^{k} a_j p^j, \quad 0 \leq a_j < p,$$

is the p-adic expansion of the positive integer n.

Solution by Y. H. Harris Kwong, SUNY College at Fredonia, Fredonia, NY. Let $f(r)$ denote the number of terms in the expansion of $(x_1 + x_2 + \cdots + x_m)^r$ whose coefficients are not divisible by p. For $0 \leq r < p$, since none of the terms in the expansion of $(x_1 + x_2 + \cdots + x_m)^r$ has coefficient divisible by p, $f(r)$ is precisely the total number of terms in the expansion. Thus $f(r)$ equals the number of combinations of $\{x_1, x_2, \ldots, x_m\}$ taken r at a time, with repetitions allowed. This is

known to be $\binom{r+m-1}{m-1}$. Now from

$$(x_1 + x_2 + \cdots + x_m)^n = \prod_{i=0}^{k} (x_1 + x_2 + \cdots + x_m)^{a_i p^i}$$
$$\equiv \prod_{i=0}^{k} \left(x_1^{p^i} + x_2^{p^i} + \cdots + x_m^{p^i}\right)^{a_i} \pmod{p},$$

it is evident that

$$f(n) = \prod_{i=0}^{k} f(a_i) = \prod_{i=0}^{k} \binom{a_i + m - 1}{m - 1}.$$

Editorial comment. As pointed out by several solvers, this result can be found in the paper of F. T. Howard, "The number of multinomial coefficients divisible by a fixed power of a prime," Pacific J. Math. 50 (1974), 99–108 and in the solution by David Singmaster of Problem 1255, *Mathematics Magazine* 61 (1988) 53.

Solved also by R. Betts (student), J. C. Binz (Switzerland), D. Callan, R. J. Chapman (United Kingdom), B. de Weger (The Netherlands), F. Dodd, F. T. Howard, J. Heuver (Canada), H. Ki (Korea), H. K. Krishnapriyan, E. Levine, O. P. Lossers (The Netherlands), J. H. Nieto (Venezuela), A. Nijenhuis, A. Pedersen (Denmark), J. H. Steelman, C. Wildhagen (The Netherlands), D. Zeilberger, A. Zulauf (New Zealand), the Stanford Ordinal, and the proposer.

An Inequality for the Brocard Angle

E 3314 [1988, 253]. *Proposed by Antal Bege, Miercurea Ciuc, Romania.*

Let M be a point inside an acute-angled triangle ABC, and put $A_1 = \measuredangle MAC$, $B_1 = \measuredangle MBA$, $C_1 = \measuredangle MCB$. Prove that

$$\cot A_1 + \cot B_1 + \cot C_1 > 2^{3/2}3^{1/4}(\cot A + \cot B + \cot C)^{1/2}.$$

Solution by Marcin E. Kuczma, University of Warsaw, Warsaw, Poland. We prove that

$$\cot A_1 + \cot B_1 + \cot C_1 \geqslant 3^{5/4}(\cot A + \cot B + \cot C)^{1/2}, \tag{1}$$

which implies the desired inequality and holds with equality when the triangle is equilateral and M is its center.

We have $MA/MB = \sin(B_1)/\sin(A - A_1)$ and analogously for MB/MC and MC/MA, which yields

$$\frac{\sin(A - A_1)\sin(B - B_1)\sin(C - C_1)}{\sin(A_1)\sin(B_1)\sin(C_1)} = 1. \tag{2}$$

(This is a form of Ceva's Theorem.) Using a, b, c and x, y, z for the cotangents of A, B, C and A_1, B_1, C_1, respectively, the well-known identity $\tan A + \tan B + \tan C = \tan A \tan B \tan C$ becomes

$$a^{-1} + b^{-1} + c^{-1} = (abc)^{-1}. \tag{3}$$

Furthermore, we have $\sin(A - A_1)/\sin A_1 = \sin A(\cot A_1 - \cot A) = (1 + a^2)^{-1/2}(x - a)$, and similarly for b, c, which transforms (2) into

$$(x - a)(y - b)(z - c) = \left[(1 + a^2)(1 + b^2)(1 + c^2)\right]^{1/2}. \tag{4}$$

This reduces the problem to minimizing the quantity

$$Q = (x + y + z)/(a + b + c)^{1/2}$$

subject to the equalities (3), (4) and the inequalities $0 < a < x$, $0 < b < y$, and $0 < c < z$ that hold because the triangle is acute-angled. By the Arithmetic-Geometric Mean Inequality, we have $(x - a) + (y - b) + (z - c) \geqslant 3[(x - a)(y - b)(z - c)]^{1/3}$, which with (4) implies

$$x + y + z \geqslant a + b + c + 3\left[(1 + a^2)(1 + b^2)(1 + c^2)\right]^{1/6}. \tag{5}$$

Note that the Mean Inequality implies that for a *fixed* triangle, the sum $x + y + z$ attains its minimum only when M is chosen so that $x - a = y - b = z - c$, in which case the common value is $[(1 + a^2)(1 + b^2)(1 + c^2)]^{1/6}$.

Using $s = a + b + c$ and $q = abc$, we can multiply (3) by q and square it to rewrite it as $a^2b^2 + b^2c^2 + c^2a^2 = 1 - 2qs$. Further (3) implies $a^2 + b^2 + c^2 = (a + b + c)^2 - 2ab - 2bc - 2ca = s^2 - 2$, and together these two equations imply

$$(1 + a^2)(1 + b^2)(1 + c^2) = 1 + (s^2 - 2) + (1 - 2qs) + q^2 = (s - q)^2. \tag{6}$$

Now (3) and Cauchy's inequality give $s/q = (a + b + c)(a^{-1} + b^{-1} + c^{-1}) \geqslant 9$, which implies $s - q \geqslant 8s/9$, with equality if and only if $a = b = c$. Together with (5), (6), this produces the following inequality:

$$Q \geqslant s^{-1/2}\left(s + 3(s - q)^{1/3}\right) \geqslant s^{-1/2}\left(s + 3(8s/9)^{1/3}\right) = s^{1/2} + 2 \cdot 3^{1/3}s^{-1/6}$$

Let $\varphi(s) = s^{1/2} + 2 \cdot 3^{1/3}s^{-1/6}$.

Using $a^2 + b^2 + c^2 = s^2 - 2$ and Cauchy's inequality, we have

$$s = a + b + c \leqslant 3^{1/2}(a^2 + b^2 + c^2)^{1/2} = (3s^2 - 6)^{1/2},$$

which implies $s \geqslant 3^{1/2}$. Routine computation by calculus shows that φ increases in $[3^{1/2}, \infty)$. Thus, finally, $Q \geqslant \varphi(3^{1/2}) = 3^{5/4}$, which implies (1). All inequalities convert into equalities for the equilateral triangle and its center, which is the only case where they do so.

Editorial comment. If ω is the Brocard angle of ABC (see R. A. Johnson, *Modern Geometry*, Houghton-Mifflin (New York, 1929)), then $\cot \omega = \cot A + \cot B + \cot C$, and the improved inequality (1) is equivalent to $\cot A_1 + \cot B_1 + \cot C_1 \geqslant 3(\sqrt{3} \cot \omega)^{1/2}$.

The improved inequality (1) was obtained also by Jean Anglesio (France), Eugene Levine, and Harvey Lipman. Levine remarked that in fact (1) holds for all triangles, not just for acute-angled triangles. Walther Janous (Austria) proved the original assertion of the problem by showing that

$$\cot A_1 + \cot B_1 + \cot C_1 \geqslant \cot A + \cot B + \cot C + 2\sqrt{3}$$
$$\geqslant 2^{3/2}3^{1/4}(\cot A + \cot B + \cot C)^{1/2}.$$

One incorrect solution was received.

The Evolute of an Ellipse

E 3316 [1989, 254]. *Proposed by Jim Delany, California Polytechnic State University, San Luis Obispo, CA.*

Consider the fluctuations in distance between an observer located at (h, k) and an object orbiting along the ellipse $x^2/a^2 + y^2/b^2 = 1$, where $a > b > 0$. If the

observer is at or near the origin, then on each orbit the distance will have two relative maxima and two relative minima. If the observer is far from the origin, there will be only one of each. Determine the equation of the curve that marks the boundary between the two cases.

Solution by K. F. Andersen, University of Alberta, Edmonton, Alberta. The desired equation is $(ah)^{2/3} + (bk)^{2/3} = (a^2 - b^2)^{2/3}$. Let $f(t) = (a\cos t - h)^2 + (b\sin t - k)^2$ represent the square of the distance from (h, k) to the point $(a\cos t, b\sin t)$ on the ellipse. By Rolle's Theorem, there is a zero of f'' between each pair of consecutive zeros of f'. Thus the boundary curve is characterized by the requirement that there exists a value of t such that $f'(t) = f''(t) = 0$, that is,

$$(b^2 - a^2)\sin t\cos t + ah\sin t - bk\cos t = 0,$$
$$(b^2 - a^2)(\cos^2 t - \sin^2 t) + ah\cos t + bk\sin t = 0.$$

Multiplying the first by $\sin t$, the second by $\cos t$, and adding yields $\cos^3 t = ah/(a^2 - b^2)$. Multiplying the first by $\cos t$, the second by $\sin t$, and subtracting yields $\sin^3 t = -bk/(a^2 - b^2)$. We combine these using the identity $\sin^2 t + \cos^2 t = 1$ to obtain the equation claimed.

Editorial comment. Many readers noted that this equation describes a concave four-pointed star, known as an astroid. The astroid is well-known to be the evolute of the ellipse, defined to be the locus of the centers of its inscribed circles of curvature. The astroid is also the locus of points at which three normals to the curve meet (except for the cusps), separating the region of the plane where four normals meet from that where only two meet. Finally, the astroid is also known to be the envelope of the normals to the ellipse. G. D. Chakerian noted that this is a very old problem, essentially solved by Appolonius of Perga in his first book on Conics.

Solved also by R. Bagby, H. Carus, G. DuChamp & J.-Y. Thibon (France), J. Duemmel, A. Guetter, H. Guggenheimer, D. Halprin (Australia), M. E. Kuczma (Poland), K.-W. Lau (Hong Kong), H. Lipman, O. P. Lossers (Netherlands), L. F. S. Martins, J. McHugh, R. B. Nelsen, A. Nijenhuis, T. S. Norfolk, R. Parris, A. J. Schwenk, J. V. Wales Jr., Western Maryland College Problems Group, and the proposer.

Dini Rides Again

E 3318 [1989, 254]. *Proposed by Walter Rudin, University of Wisconsin, Madison.*

Suppose $0 < p < \infty$. Let A be the set of sequences of positive real numbers $a_1, a_2, \ldots$ such that $\sum_{n=1}^{\infty} a_n = 1$. Find $\sup_A \sum_{n=1}^{\infty} a_n(a_n + a_{n+1} + \cdots)^{p-1}$ and $\sup_A \sum_{n=1}^{\infty} a_n(a_1 + a_2 + \cdots + a_n)^{p-1}$.

Solution by Albert Nijenhuis, Seattle, WA. Both suprema are 1 when $p \geqslant 1$ but $1/p$ when $0 < p \leqslant 1$.

Instead of determining the suprema of the infinite series

$$\sum_{n=1}^{\infty} a_n(a_1 + \cdots + a_n)^{p-1} \quad \text{and} \quad \sum_{n=1}^{\infty} a_n(a_n + a_{n+1} + \cdots)^{p-1}$$

we observe that these are the same as the suprema of the finite sums of the form

$$\sum_{n=1}^{N} a_n(a_1 + \cdots + a_n)^{p-1} \quad \text{and} \quad \sum_{n=1}^{N} a_n(a_n + \cdots + a_N)^{p-1},$$

with unbounded N, $a_n > 0$ and $\Sigma_{n=1}^{N} a_n = 1$.

The first sum is a "right endpoint" Riemann sum for $\int_0^1 x^{p-1}\,dx$. If $p > 1$ then $x \mapsto x^{p-1}$ is increasing in $[0, 1]$, and the sum is an upper estimate for the integral. The largest "right endpoint" sum consists of one term ($N = 1$), $1^{p-1} \cdot 1$, yielding the value 1. Of course, this can be arbitrarily closely approximated by an infinite series with a_1 slightly less than 1, and the other positive a_n making up the balance. If $0 < p \leqslant 1$, then $x \mapsto x^{p-1}$ is decreasing (non-increasing), and the "right endpoint" Riemann sum is a lower estimate of $\int_0^1 x^{p-1}\,dx = 1/p$, so that $1/p$ is the supremum.

The second sum, which can be written as

$$\sum_{n=1}^{N} a_n\{1 - (a_1 + \cdots + a_{n-1})\}^{p-1},$$

is a "left endpoint" Riemann sum for the integral $\int_0^1 (1 - x)^{p-1}\,dx$. When $p > 1$, $x \mapsto (1 - x)^{p-1}$ is a decreasing function, and the sum is an overestimate. Again, the one term sum $(1 - 0)^{p-1} \cdot 1 = 1$ is the supremum. If $0 < p \leqslant 1$, the sum is again a lower estimate, and the supremum equals the integral, i.e., $1/p$.

Editorial comment. David Callan noted that the problem is essentially the same as E 2996 [1983, 334; 1986, 303]. According to §39 of Knopp's *Infinite Series*, Blackie, 1928, the assertion of the problem essentially goes back to Dini.

Solved also by K. F. Andersen (Canada), L. Baribeau (Canada), D. Borwein (Canada), D. Callan, G. Duchamp (France), M. Falkowitz (Israel), J. Gaisser, E. Hertz, H. Ki (Korea), M. E. Kuczma (Poland), K.-W. Lau (Hong Kong), O. P. Lossers (The Netherlands), K. Schilling, M. Stieglitz and H. Tietz (West Germany), Western Maryland College Problem Group, and the proposer.

Rationally Independent Subsets of the Reals

E 3319 [1989, 356]. *Proposed by Paul Erdös, Hungarian Academy of Science, and János Surányi, Eötvös Loránd University, Budapest, Hungary.*

Suppose S is a set of $2n + 1$ irrational real numbers. Prove that S has a subset T of cardinality $n + 1$ such that no nonempty subset of T has a rational sum.

Solution I by Ilias G. Kastanas, California State University, Los Angeles. Suppose $S = \{\alpha_0, \alpha_1, \ldots, \alpha_{2n}\}$. Let $1, \beta_1, \beta_2, \ldots, \beta_m$ be a basis for the finite-dimensional vector space V over $\mathbb{Q}$ generated by $1, \alpha_0, \alpha_1, \ldots, \alpha_{2n}$. (Here $\mathbb{Q}$ denotes the field of rational numbers.) Let $p(\alpha_i)$ be the projection of α_i on the subspace of V complementary to 1, i.e., let $\alpha_i = r_i + p(\alpha_i)$, where r_i is rational and $p(\alpha_i)$ is a linear combination of $\beta_1, \ldots, \beta_m$ with rational coefficients not all zero ($i = 0, 1, \ldots, 2n$). Then any set of α_i has a rational sum if and only if the corresponding set of $p(\alpha_i)$ has sum zero. Hence it suffices to show that there exists a set of $n + 1$ of the $p(\alpha_i)$ such that each of its non-empty subsets has a non-zero sum; but this is immediate, since either $n + 1$ of the $p(\alpha_i)$ are positive or $n + 1$ of the $p(\alpha_i)$ are negative.

In the assertion of the problem the number $n + 1$ cannot be replaced by a larger one, as is shown by the example $\alpha_i = (-1)^i e + i$ for $i = 0, 1, \ldots, 2n$.

Solution II by an anonymous contributor. Let W be the finite-dimensional vector space over $\mathbb{Q}$ generated by the elements of S. Since S is a finite set of nonzero elements in W, there exists a hyperplane through the origin of W containing no elements of S. Clearly $n + 1$ of the elements of S fall on the same side of this hyperplane. These $n + 1$ elements of S form a set T having the property stated in the problem; in fact no nontrivial nonnegative linear combination of the elements of T equals zero.

Solved also by I. C. Bivens, R. J. Chapman (England), L. E. Clarke (England), J. Ferrer (Spain), S. Gaignoux (France), J. Henle, I. M. Isaacs, O. P. Lossers (The Netherlands), H. Noland, A. Riese, P. Shirts, P. A. Staley, J. H. Steelman, R. Stong, O. Šuch (Czechoslovakia), P. Tracy, E. Triesch (West Germany), T. P. Vaughan, and the proposers.

To Be or Not to Be Injective

E 3325 [1989, 445]. *Proposed by Walter Rudin, University of Wisconsin, Madison.*

Let us say that a function f of the form

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n$$

has property P_t if $\sum_{n=2}^{\infty} n|a_n| \leq t$. Prove:

a) If P_t holds for some $t < \infty$, then f is continuous on the closed unit disc, i.e. on $\{z \in \mathbb{C} | z| \leq 1\}$.

b) If P_1 holds, then f is one-to-one on the closed unit disc.

c) If $t > 1$, then some f satisfies P_t without being one-to-one in the open unit disc.

Solution to (a) by Raymond Mortini, Universität Karlsruhe, Karlsruhe, West Germany. If f satisfies property t for some $t > 0$, then the power series for f' and that for f converge uniformly (and absolutely) on $\overline{D} = \{z: |z| \leq 1\}$. Hence f and f' have continuous extensions to $\overline{D}$.

Solution to (b) by Adam Riese, Wright State University, Dayton, OH. If P_1 holds for f, choose $z, w \in \overline{D}$ with $z \neq w$. Then $|z + w| < 2$ and $|z^{n-1} + z^{n-2}w| < 2$. Thus $|z^n - w^n| = |z - w||z^{n-1} + z^{n-2}w + \cdots + w^{n-1}| < |z - w|n$. If $f(z) \not\equiv z$, then $|f(z) - f(w)| \geq |z - w| - |\sum_{n=2}^{\infty} a_n(z^n - w^n)| > |z - w|(1 - \sum_{n=2}^{\infty} |a_n|n) \geq 0$. Hence f is one-to-one.

Solution to (c) by K. F. Andersen, University of Alberta, Edmonton, Alberta. Given $t > 1$, choose $b > 0$ such that $(1/t)^2 + b^2 < 1$. Set $f(z) = z + (t/2)z^2 = (t/2)(z + 1/t)^2 - 1/(2t)$; note that f has property P_t. For $z_1 = -t^{-1} + bi$ and $z_2 = -t^{-1} - bi$, we have $|z_1| = |z_2| < 1$ and $f(z_1) = -(1 + b^2t^2)/(2t) = f(z_2)$, so f is not one-to-one in $\{z: |z| < 1\}$.

Editorial comment. Raymond Mortini remarked that parts (b) and (c) are Problem 24(a) on p. 73 of P. L. Duren's *Univalent Functions*. A. W. Goodman remarked that under the hypothesis of (b) the function f maps the closed unit disc

onto a domain that is starlike with respect to the origin; see *Proc. Amer. Math. Soc.* 8 (1957) 598–601.

Solved by the proposer and 23 others. Five incorrect solutions were received.

Lines Normal to a Cubic at Two Points

E 3334 [1989, 524]. *Proposed by Stanley Rabinowitz, Westford MA.*

Consider the cubic curve $y = x^3 + ax^2 + bx + c$, where a, b, c are real numbers with $a^2 - 3b > 3\sqrt{3}$. Prove that there are exactly two lines that are perpendicular (normal) to the cubic at two points of intersection and that these two lines intersect at the point of inflection of the curve.

Solution by Luiz Felipe Martins, Brown University, Providence, RI. We may assume $a = c = 0$; the general case is reduced to this by considering a coordinate translation that places the origin at the point of inflection of the cubic. The slope of the line tangent to the cubic at the inflection point is $b - a^2/3$, so the quantity $a^2 - 3b = -3(b - a^2/3)$ is invariant under translation.

Hence we assume the curve is $y = f(x) = x^3 + bx$ with $b < -\sqrt{3}$. Let $P = (p, f(p))$ and $Q = (q, f(q))$ be points on the curve such that PQ is normal to the curve at P and Q. The tangents at P and Q must be parallel, so $f'(p) = f'(q)$. Since f' is an even quadratic function, this implies $q = -p$. Since f is odd, this implies that PQ contains the origin, so all such "double normals" intersect there.

To determine how many such PQ there are, we may assume $p > 0$. With $q = -p$, the slope of PQ is $f(p)/p$, so PQ is normal to the curve at p and q if and only if $f'(p) = -p/f(p)$, or $3p^4 + 4bp^2 + b^2 + 1 = 0$. The solutions of this quadratic in p^2 are $p^2 = (-2b \pm \sqrt{b^2 - 3})/3$. With $b < -\sqrt{3}$, both solutions are real and positive, and the positive square roots of these are the desired values of p. These yield exactly two distinct solutions PQ.

Solution II by Michael Vowe, Therwil, Switzerland. As above, we reduce the problem to $f(x) = x^3 + bx$, $b < -\sqrt{3}$. Let $y = mx + d$ be a line normal to the curve at $(p, f(p))$ and $(q, f(q))$. As above, $f'(p) = f'(q)$, which with normality implies $3p^2 + b = -1/m = 3q^2 + b$. This yields $q = -p$, $f(q) = -f(p)$, which implies $m = f(p)/p = p^2 + b$, $d = 0$. Hence $-1/m = f'(p) = 3p^2 + b = 3m - 2b$. The solutions to this quadratic in m are $m = (b \pm \sqrt{b^2 - 3})/3$, both negative for $b < -\sqrt{3}$. Since $d = 0$, the two lines intersect at $(0, 0)$, the point of inflection.

Solved by 35 readers and the proposer. One incomplete solution was received.

A Dense Set Determined by a Polynomial

E 3336 [1989, 525]. *Proposed by Nick MacKinnon, Winchester, England.*

Suppose P is a given polynomial with integer coefficients and degree greater than one. Let

$$W = \{w: w \text{ is an irrational real number but } P(w) \text{ is rational}\}.$$

Prove that W is dense on the real line.

Solution by Stewart A. Levin, Mobil Research and Development, Dallas, TX. We exhibit a set S of rational numbers that is dense and disjoint from $P(\mathbb{Q})$. Since P is continuously differentiable and has a finite number of isolated extrema, the real line can be decomposed into a finite number of intervals on each of which P is monotonic. By considering each such interval separately, it is easy to see that the set of irrational w having $P(w)$ in such a set is dense on the real line. With $P(w) = \sum_{j=0}^{n} a_j w^j$, let S be the set of rational numbers p/q such that q is a prime not dividing p or a_n. If $p/q = P(w)$ for $p/q \in S$ and $w \in \mathbb{Q}$, then $w = r/(sq)$ for some integers r, s with $\gcd(q, r) = 1$. For if z is a rational number with denominator relatively prime to q, then the denominator of the rational number $P(z)$ is also relatively prime to q. However, $P(r/(sq)) = \sum_{j=0}^{n} a_j r^j (sq)^{n-j}/(sq)^n$. Since q divides all the terms in the numerator except $a_n r^n$, the numerator is relatively prime to q. Hence the denominator of the reduced fraction equal to $P(r/(sq))$ contains a factor of q^n with $n > 1$, contradicting $P(w) = p/q$.

Editorial comment. The proposer and several other solvers used the Eisenstein Irreducibility Criterion to prove that there is a dense set of rational numbers a such that $P(x) - a$ is irreducible over $\mathbb{Q}$. This implies that there is a dense set of algebraic numbers w of *exact degree* n such that $P(w)$ is rational.

Solved also by 26 other readers and the proposer.

ADVANCED PROBLEMS

6640. *Proposed by Douglas Bowman (student), UCLA.*

Let $f(n)$ be the number of partitions of the positive integer n into parts taken from the set of divisors of n. Prove that

$$\{1 + o(1)\}\left\{\frac{\tau(n)}{2} - 1\right\}\log n \leqslant \log f(n) \leqslant \{1 + o(1)\}\frac{\tau(n)}{2}\log n,$$

where $\tau(n)$ is the number of divisors of n.

6641. *Proposed by Theodore M. Alper, Stanford University, Stanford, CA.*

(a) For every positive integer n, is there an integer N_n such that every finite group of order at least N_n has an abelian subgroup of order at least n?

(b) For every positive integer n, does every infinite group have an abelian subgroup (possibly infinite) of order at least n?

6642. *Proposed by the editors.*

Let λ be the maximum possible inradius of an arbitrary triangle lying in the closed set bounded by a regular pentagon of side-length one.

(a)* Determine λ up to an error of at most 10^{-3}.

(b)* Determine λ exactly.

Cf. 6477 [1984, 588; 1989, 945] and 6478 [1984, 588; 1990, just below].

SOLUTIONS OF ADVANCED PROBLEMS

Inradii and Width

6478 [1984, 588]. *Proposed by Louis Funar, University of Craiova, Rumania.*

Let r be the radius of the incircle of an arbitrary triangle lying in a closed convex figure F of width w, and let R be the radius of the incircle of F. Are the following inequalities valid?

(i)* $\frac{1}{4} \leqslant (\sup r)/w \leqslant \frac{1}{2}$,

(ii)* $\frac{1}{2} \leqslant (\sup r)/R \leqslant 1$.

Solution (partial) by C. A. Rogers, University College, London. Three of the inequalities are easy; the remaining one is, we believe, invalid.

Since the width of the incircle of the triangle is $2r$, we clearly have $2r \leqslant w$.

Since the incircle of the triangle is contained in F, we clearly have $r \leqslant R$.

Since the incircle of F contains an equilateral triangle with incircle of radius $\frac{1}{2}R$, we clearly have $(\sup r)/R \geqslant \frac{1}{2}$.

The remaining inequality $(\sup r)/w \geqslant \frac{1}{4}$ appears to be incorrect. To see this consider the special case when F is a regular pentagon $ABCDE$. We work on the *assumption*, which seems highly plausible, that the triangles, for which the largest inradius is attained, are isosocles triangles inscribed in the pentagon, the vertex where the equal sides meet being at one of the vertices, say A, of the pentagon. Let the second and third vertices of the triangle be F on CB and G on DE, the ratios CF/CB and DG/DE having a common value t, with $0 \leqslant t \leqslant 1$. Simple numerical calculations show that the maximum value for r/w is attained for t near 0.27 and is less than 0.2441. Thus, provided our assumption is justified, we do not have $(\sup r)/w \geqslant \frac{1}{4}$ when F is a regular pentagon.

Editorial comment. Although the inequality $(\sup r)/w \geqslant \frac{1}{4}$ seems to be incorrect, it is true that $(\sup r)/w$ has a positive lower bound independent of F. For by Theorem 50 of H. G. Eggleston, *Convexity*, Cambridge Tract No. 47, 1958, we have $R \geqslant w/3$; since $\sup r \geqslant R/2$, it follows that $(\sup r)/w \geqslant \frac{1}{6}$. (This weaker inequality was noted by the proposer.)

To complete Rogers' argument that the inequality $(\sup r)/w \geqslant \frac{1}{4}$ is not always true, one needs a rather precise determination of the maximal possible inradius of a triangle lying in the closed set bounded by a regular pentagon of width one, or at least a proof that this maximum is strictly less than $\frac{1}{4}$. Accordingly we have posed this question as Advanced Problem 6642 above. (Since 6642 is stated in terms of side-length, we note that for a regular pentagon

$$\frac{\text{width}}{\text{side-length}} = \frac{1 + \cos(\pi/5)}{2\cos(\pi/5)} = \frac{5 + \sqrt{5}}{2\sqrt{10 - 2\sqrt{5}}} = 1.5388418.\Big)$$

It remains to find the greatest lower bound of $(\sup r)/w$ as F ranges over *all* convex figures of width one and to determine whether or not the convex set bounded by a regular pentagon is the extreme case. We hope that some of our readers will consider this question as well.

The only solution received before the customary deadline was by L. Kuipers (Switzerland), who gave a proof of the inequalities of the problem for the special case when F is a parallelogram. The solution given above was prepared by Professor Rogers at the request of the editors.

The Fejér-Jackson Sum and Legendre Polynomials

6585 [1988, 880]. *Proposed by Alexandru Lupaş, Sibiu, Romania.*

Prove that if $-1 < x < 1$, then

$$\sum_{k=1}^{n} \frac{\sin(k \arccos x)}{k} = \frac{\sqrt{1-x}}{2} \int_{-1}^{x} \frac{1 - P_n(y)}{1-y} \frac{dy}{\sqrt{x-y}},$$

where

$$P_n(x) = \frac{1}{2^n n!} \left(\frac{d}{dx}\right)^n (x^2 - 1)^n$$

is the nth Legendre polynomial.

Solution by the Western Maryland College Problems Group, Westminster, Maryland. We will use induction. The case $n = 1$ is easily checked. If the assertion is true up to $n - 1$, the case for n reduces to showing that

$$\frac{1}{2}(1+x)^{-1/2} \int_{-1}^{x} n(P_{n-1}(t) - P_n(t))(1-t)^{-1}(x-t)^{-1/2}\, dt = (1-x^2)^{-1/2} \sin(n \arccos x), \tag{1}$$

i.e., to showing that $L_{n-1}(x) = U_{n-1}(x)$, where $L_{n-1}(x)$ is the left-hand side of (1) and $U_{n-1}(x) = \sin(n \arccos x)/\sin(\arccos x)$ is the Chebyshev polynomial of the second kind. We require the Christoffel-Darboux formula for Legendre polynomials

$$n(P_{n-1}(t) - P_n(t))(1-t)^{-1} = \sum_{k=0}^{n-1} (2k+1) P_k(t). \tag{2}$$

(This formula (2) may be derived by summation from the familiar recurrence formula

$$x(2k+1)P_k(x) = (k+1)P_{k+1}(x) + kP_{k-1}(x).)$$

If we make the substitution $t = x - (1+x)s^2$ and use (2), then $L_{n-1}(x)$ becomes

$$L_{n-1}(x) = \sum_{k=0}^{n-1} (2k+1) \int_0^1 P_k(x - (1+x)s^2)\, ds. \tag{3}$$

Now from the classical generating function

$$\sum_{k=0}^{\infty} P_k(t) t^{k+1/2} = t^{1/2}(1 - 2xt + t^2)^{-1/2},$$

we obtain by differentiation

$$\sum_{k=0}^{\infty} (2k+1) P_k(t) t^k = (1-t^2)(1-2xt+t^2)^{-3/2}. \tag{4}$$

Multiplying (3) by t^{n-1}, summing from $n = 1$ to ∞, changing the order of

summation, and using (4), we obtain

$$\sum_{n=0}^{\infty} t^{n-1}L_{n-1}(x) = \sum_{k=0}^{\infty}(2k+1)\sum_{n=k+1}^{\infty} t^{n-1}\int_0^1 P_k\big(x-(1+x)s^2\big)\,ds$$

$$= \int_0^1 (1-t)^{-1}\sum_{k=0}^{\infty}(2k+1)P_k\big(x-(1+x)s^2\big)t^k\,ds$$

$$= (1+t)\int_0^1 \{1-2xt+t^2+2t(1+x)s^2\}^{-3/2}\,ds$$

$$= (1-2xt+t^2)^{-1}$$

by an elementary integration. Since

$$\sum_{n=0}^{\infty} t^{n-1}U_{n-1}(x) = (1-2xt+t^2)^{-1},$$

(1) follows.

Editorial comment. Most of our solvers were rather cavalier about interchange of limit processes, but David Callan provided a proof somewhat similar to the above that justified such operations by uniform convergence and the Moore-Osgood theorem. Some solvers demonstrated a knowledge of Jacobi polynomials, while others relied on the formula

$$\int_{-1}^{x}\frac{P_n(t)\,dt}{\sqrt{x-t}} = \frac{2}{(2n+1)\sqrt{1+x}}\big(T_n(x)+T_{n+1}(x)\big),$$

where $T_n(x) = \cos(n \arccos x)$ is the Chebyshev polynomial of the first kind (see, e.g., Abramowitz and Stegun, *Handbook of Mathematical Functions*, Formula 22.13.10, p. 786).

Since $|P_n(t)| < 1$ for $-1 < t < 1$, this provides, as the proposer remarks, a proof of the Fejér-Jackson inequality

$$\sum_{k=1}^{n}\frac{\sin k\theta}{k} > 0, \qquad 0 < \theta < \pi.$$

Richard Askey informs us that despite about 20 previous proofs, this one appears new! For a proof of Turán's inequality by a similar method, see 6517 [1986, 305; 1988, 264].

Solved also by Jean Anglesio (France), James P. Butler, Paul F. Byrd, David Callan, Louis N. Howard, A. A. Jagers (The Netherlands), Haseo Ki (Korea), Allen R. Miller, I. A. Sakmar (Turkey), and the proposer.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

Learning Discrete Mathematics With ISETL. By Nancy Baxter, Ed Dubinsky, and Gary Levin, Springer-Verlag, 1989, xvii + 416 pp.

DONALD L. MUENCH
Department of Mathematics and Computer Science, St. John Fisher College, Rochester, New York 14618

Next to the blizzard of lectures, workshops, minicourses and publicity surrounding calculus reform, there is a brisk breeze surrounding discrete mathematics. Starting with Anthony Ralston's controversial article [**6**] on the role of discrete mathematics in the curriculum and its threat to the privileged status of calculus as the pearly gateway to mathematical exaltation, there have been as many ideas about discrete mathematics as there are mathematicians. But, if you examine the textbooks on the subject, there are more similarities than differences, although there are books that approach the subject in radically different ways. Fred Roberts' *Discrete Mathematical Models* [**7**] is one significant example. And, the book under review, *Learning Discrete Mathematics with ISETL*, by Nancy Baxter, Ed Dubinsky, and Gary Levin, is yet another noteworthy exception. This book stands apart from the others in its unique approach, in its grounding in a distinctive theory of learning, and in its use of a not-so-familiar (but steadily becoming better known) computer language as a powerful medium of mathematical expression.

As soon as one hears or sees the words "computer" and "mathematics" in close proximity, the reaction is either enthusiasm or alarm. The alarming unease is often accompanied by a sense of sadness—as though something valuable has been lost or defiled—and a concern that some important topic in the curriculum will be omitted or diminished in order to make room for "the programming".

In contrast, the enthusiasm derives from the conviction that the use of the new technology should and will radically alter the present curriculum in the direction of encouraging the kinds of mathematical skill and thinking that have either been missing or could not have been done effectively before.

However, the computer should not be the focus of the debate. The computer is not the central element in the discussion; how students learn is what we should be talking about. Using the computer can motivate, help, and reinforce student learning. The idea is to concretize the abstract concepts, to make them tangible and real to the student. Only then can our students make the mental images that breathe life into abstractions.

Because learning is central, it is significant that the title of this book is not "Discrete Mathematics with ISETL". The "Learning" in the title is a clue that this book is different and has a definite orientation toward the student. This is reason enough to read it and act on it with your (essentially) free copy of ISETL, which unfortunately does not come with the book. But, you may get ISETL and documentation from Ron Pullins of West Educational Publishing.

The approach taken by the authors is the result of basic research done by the first two authors [**1**], [**2**] on how students learn. Their theory, heavily based on the

work of Piaget, is used to develop the topics in the book, carefully constructing computer activities which concretize concepts such as propositional calculus, multilevel quantification, induction, relations, functions, graphs, and combinatorics. Their findings show that students CAN and DO learn mathematics by writing small computer programs. Anyone adopting this book needs to be aware that it calls for the active participation of both the student and the teacher. One does not read this book without sitting down at a computer terminal and DOING something with the concepts that are being explained. And the beauty of it is that no coercion is necessary; the students become the most important players and the influence of the teacher recedes into the background. They DO the work and interact with one another with very little prompting from the professor. They WANT to do it, and they ENJOY doing it. Oh, yes, they do get frustrated from time to time, but they don't give up.

Creative approaches such as this may very well revolutionize the way we teach because it emphasizes the role of the student as the primary agent of learning. Seymour Papert [**5**] has described our traditional lecture method as one which makes the students "passive observers" and not "active participants." What students often do in class is transcribe words from the instructor's tongue onto paper. Then, when they study, they employ the method of pattern-matching, hoping to find examples in the book or their notes which resemble the problem they are working on. And, when they imitate us and their textbook, we reward them. And then we complain that they really haven't learned to be independent thinkers.

So how do we remedy this situation? Because the main goal of our teaching is to facilitate learning, our teaching should be informed by the results of investigations by scholars into the way students learn. This knowledge should shape our instructional strategies and influence the way we present mathematics in our textbooks. With the proper materials, written from the point of view of how students learn, mathematics may be learned more effectively. And, if computers are in the classroom laboratory (someday "laboratory" and "classroom" will be synonymous!), there will be little lecturing and more interactive learning and discussion.

What is going on in the classroom when a computer is being used? Certainly not the copying of notes that the instructor has written on a blackboard! What is happening is active learning. By performing carefully designed computer activities, the students develop mental images which they use to construct an abstract mathematical concept. They concretize abstract ideas through computer representations. Using a language like ISETL, students learn about properties of mathematical objects by writing "property-testers." These are ISETL statements which look like the familiar descriptions and definitions in precise mathematical notation. The computer is the powerful means to the end of learning and developing mathematical maturity and sophistication. It is not the end itself.

What is ISETL? Before an answer can be given, it should be said that the language and the software should be chosen for their suitability in promoting student learning. If it is easy to learn and powerful enough to enable students to form the mental images we want them to construct, and encourages experimentation and discovery, then it may be considered in a course. ISETL is a powerful high-level language whose syntax is close to standard mathematical notation. It is easy to learn, read, and write. Derived from SETL [**8**] and developed by Gary Levin [**3**], with substantial input from the other authors, as a tool for the teaching

and learning of mathematical concepts, ISETL (Interactive SET Language) has the expressive power to construct concrete realizations of mathematical concepts. It supports boolean expressions and operations (and, or, not, implication), heterogeneous sets and tuples (sequences that have no fixed size), iteration over sets, tuples and strings, and both universal and existential quantification. ISETL also supports the usual control statements (while-do, for-do and if-then-else). There are no graphics with the ISETL package, but a new version now under development will include it.

ISETL has no procedures, only functions. Functions (funcs in ISETL) are first-class objects—that is, they may be used just like any other data object. Consequently, they can be elements of sets and sequences and can be input parameters in a function and be returned by a function. This powerful feature provides incredible flexibility to express concretely not only the idea of a function itself, but other abstract concepts as well. A more extensive review of ISETL can be found in [**4**].

Probably the most effective chapter of all in the book is the chapter on functions. This one chapter may be the best explanation and development of the idea of a function that has appeared in any textbook. Considerable effort is taken to lead the student to a thorough understanding of the ways in which functions may be represented: as a process, as a set of ordered pairs, and (for those whose domains are positive integers) strings and sequences. The ISETL exercises ask students to convert from one mode of representation to another and ask them to formulate the meaning of function composition in these different modes, thus leading them to the realization that a function is an object.

In describing a function as a set of ordered pairs, the authors say only that each element of the domain must appear exactly once as the first component of one of the pairs. The student is then asked to make the precise mathematical definition. It is this reviewer's experience that the students discover on their own several equivalent formulations of the definition. Here are some actual definitions that students have given, without the benefit of looking in some textbook (This is ISETL code, where "f" is a set of ordered pairs and $u(1)$ and $v(1)$ refer to the first components in the pairs u and v):

(1) forall x in domains (f)|forall $y1, y2$ in image (f)|
$$[x, y1] \text{ in } f \text{ and } [x, y2] \text{ in } f \text{ imp } ly1 = y2;$$
(2) forall u, v in $f|u/= v$ imp $lu(1)/= v(1)$;
(3) forall u, v in $f|u(1) = v(1)$ imp $lu = v$;
(4) $\{v(1): v \text{ in } f|(\text{exists } u \text{ in } f|\ (u/= v \text{ and } u(1) = v(1)))\} = \{\ \}$;
(5) $\{v(1) = u(1): v, u \text{ in } f|v/= u\} = \{\text{false}\}$.

Some of these look familiar, but others look strange. In all cases, they "belong" to the students who wrote them. One of the strange ones, (5), is not quite correct. A very interesting and spirited classroom discussion about (5) followed! This kind of exercise and discussion probably never would have taken place in the typical "traditional" classroom.

So, whither the computer revolution in mathematical instruction? It looks as though new and exciting materials will continue to evolve for teaching not only discrete mathematics, but also calculus, differential equations, linear algebra, and abstract algebra. But, be warned: the use of these materials may be hazardous to your former way of teaching and learning. With careful attention to the design of

appropriate activities in the classroom and in the laboratory, they just may help students learn, think, create, and grow intellectually.

REFERENCES

1. N. Baxter and E. Dubinsky, A new approach to teaching abstract mathematical concepts, *Journal of Mathematical Behavior*, to appear.
2. E. Dubinsky and P. Lewin, Reflective abstraction in mathematics education: the genetic decomposition of induction and compactness, *Journal of Mathematical Behavior*, 5 (1986) 55–92.
3. G. Levin, An Introduction to ISETL: Version 2.0, Clarkson University, 1989.
4. D. L. Muench, ISETL—interactive set language, *Notices of the American Mathematical Society*, 37 (March 1990).
5. S. Papert, Mindstorms: Children, Computers, and Powerful Ideas, Basic Books, 1980.
6. A. Ralston, Computer science, mathematics, and the undergraduate curricula in both, *American Mathematical Monthly*, 88 (1981) 472–485.
7. F. S. Roberts, Discrete Mathematical Models, Prentice-Hall, 1976.
8. J. T. Schwartz, et al., Programming with Sets: an Introduction to SETL, Springer-Verlag, 1986.

Generatingfunctionology. By Herbert S. Wilf, Academic Press, San Diego, 1990. viii + 184 pp.

RICHARD P. STANLEY
Department of Mathematics, M.I.T., Cambridge, MA 02139

Let $g(n)$ denote the number of graphs with vertex set $\{1, 2, \ldots, n\}$ (without loops or multiple edges) which are regular of degree two. Then

$$\sum_{n \geq 0} g(n) \frac{x^n}{n!} = \frac{e^{-\frac{1}{2}x - \frac{1}{4}x^2}}{\sqrt{1-x}}.$$

Who cannot be fascinated by this strange mixture of combinatorics and analysis? The theory of generating functions seeks to develop general methods for proving such identities and for using these identities to obtain further information about the coefficients (here $g(n)$). All introductory texts in combinatorics have some rudimentary information about generating functions, but Herb Wilf's *Generatingfunctionology* is the first book suitable for undergraduates to be devoted exclusively to this topic. It performs an admirable job of conveying the essential features of generating functions.

Three types of generating functions are considered in *Generatingfunctionology*: (a) ordinary generating functions, (b) exponential generating functions, and (briefly) (c) Dirichlet series. These types have the respective forms

$$\sum_{n \geq 0} f(n) x^n, \quad \sum_{n \geq 0} f(n) \frac{x^n}{n!}, \quad \sum_{n \geq 1} f(n) n^{-s}.$$

In all cases it is important to regard these series as *formal* objects, without regard to questions of convergence. If a generating function *does* converge in a suitable region of the complex plane, then techniques from complex variable theory can often be used to obtain further information about the coefficients. But even without convergence the generating function can provide valuable understanding of the coefficients. These issues are treated well in *Generatingfunctionology*.

A basic question which faces anyone wishing to investigate a sequence by the use of generating functions is the following: Which generating function should be

used? Many general situations appropriate to one type of generating function are discussed in *Generatingfunctionology*. In particular, the use of exponential generating functions in situations where one has a structure on a set (or "labelled structure") built up from "connected" pieces is explained using a colorful terminology due to Adriano Garsia. For instance, the number $f(n)$ of involutions (permutations whose square is the identity) in the symmetric group S_n is given by

$$\sum_{n\geqslant 0} f(n)\frac{x^n}{n!} = e^{x+x^2/2}.$$

The exponent $x + x^2/2$ arises because an involution is built up from 1-cycles (fixed points) and 2-cycles (transpositions). A result known as the Exponential Formula provides a general framework for understanding such results.

A nice overview is given in Chapter 4 of the basic applications of generating functions. These include finding averages and standard deviations, proving combinatorial identities and congruences, and proving results on unimodality, convexity, etc. The treatment of combinatorial identities is especially lucid and takes the reader to the frontiers of current research by providing an introduction to the WZ-method of Wilf and Zeilberger. As an example of a combinatorial identity discussed in *Generatingfunctionology*, consider the sum

$$f_n = \sum_{k=0}^{n} \binom{n+k}{2k} 2^{n-k} \qquad (n \geqslant 0).$$

If we form the generating function $y = \sum_{n\geqslant 0} f_n x^n$, then by straightforward manipulations involving little real thought we obtain

$$y = \frac{2}{3(1-4x)} + \frac{1}{3(1-x)}.$$

It follows that $f_n = (2^{2n+1} + 1)/3$. This "mechanical" procedure for evaluating combinatorial sums is analogous to the use of Laplace transforms to solve certain classes of differential equations.

The final chapter of *Generatingfunctionology* is devoted to analytic and asymptotic techniques, and will be accessible only to students who have had a course in complex variables. The three sections on asymptotics provide a brief but useful introduction to the connections between analyticity and asymptotics. There is also a short section on one of the most interesting and elegant formulas in enumerative combinatorics—the Lagrange Inversion Formula, which gives a formula for the coefficients of the generating function $f(u(t))$, where f is an arbitrary generating function, and where for a given generating function $\phi(t)$, we have that $u(t)$ is defined by the functional equation $u = t\phi(u)$. Such generating functions are common in enumerative combinatorics, and the Lagrange Inversion Formula has wide applicability, especially in situations dealing with trees and similar recursive structures. For instance, if $r(n)$ denotes the number of labeled (rooted) trees on n vertices and we set

$$u(t) = \sum_{n\geqslant 1} r(n)\frac{t^n}{n!},$$

then a simple use of the Exponential Formula yields $u(t) = te^{u(t)}$. Lagrange inversion then produces the miraculous formula $r(n) = n^{n-2}$. (There are now many other known proofs of this formula, which goes back to Borchardt, Sylvester, and Cayley.) Similarly, if $a(n)$ denotes the number of ways to divide a convex

polygon with $2n$ sides into $n-1$ quadrilaterals by drawing $n-2$ non-intersecting diagonals, then the generating function $u(t) = \sum_{n \geqslant 0} a(n) t^{2n+1}$ satisfies $u = t/(1-u^2)$, and a simple application of Lagrange inversion yields the equally miraculous result

$$a(n) = \frac{1}{2n+1}\binom{3n}{n}.$$

The treatment of Lagrange inversion in *Generatingfunctionology* is probably the weakest part of the book. While the proof Wilf provides is short and elegant, it uses the residue theorem and is only valid for suitably convergent power series. On the other hand, a proof due to Jacobi is equally short and elegant but is completely formal and uses no deep machinery. There are also purely combinatorial proofs due to Raney and to Joyal-Labelle which provide great insight into the combinatorics of trees, lattice paths, and formal languages. While the proof and applications of Lagrange inversion in *Generatingfunctionology* are disappointing, this is a small quibble about an overall quite engaging book.

What more can be done with generating functions beyond the subject matter of *Generatingfunctionology*? To whet the appetite of the reader for further study of generating functions, consider the following three enumerative functions: Let $f(n)$ be the number of ways to write the positive integer n as an ordered sum of positive integers. Let $g(n)$ be the number of ways to write an n-element set as an ordered union of pairwise disjoint nonempty subsets. Let $h(n)$ be the number of ways to write the integer $n > 1$ as an ordered product of integers > 1. Define the generating functions

$$F(x) = 1 + \sum_{n \geqslant 1} f(n) x^n, \quad G(x) = 1 + \sum_{n \geqslant 1} g(n) \frac{x^n}{n!}, \quad H(x) = 1 + \sum_{n \geqslant 2} h(n) n^{-s}.$$

Then we have

$$F(x) = \frac{1}{2 - \dfrac{1}{1-x}}, \qquad G(x) = \frac{1}{2 - e^x}, \qquad H(x) = \frac{1}{2 - \zeta(s)},$$

where

$$\frac{1}{1-x} = \sum_{n \geqslant 0} x^n, \qquad e^x = \sum_{n \geqslant 0} \frac{x^n}{n!}, \qquad \zeta(s) = \sum_{n \geqslant 1} n^{-s}.$$

Surely these formulas can't be a coincidence! There is an obvious necessity to develop a unifying theory which will "explain" the various types of generating functions and the situations to which they can be applied. Indeed, there are several such theories, based on such concepts as prefabs, incidence algebras, dissects, linked sets, and species. The various unifying theories of generating functions comprise one of many topics for study after completion of *Generatingfunctionology*. There are also a vast number of other uses and applications of generating functions, involving, for instance, the transfer-matrix method, elliptic functions, symmetric functions, Hilbert functions, convex polytopes, partially ordered sets, formal languages, lattice paths, ... Enjoy!

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, S*, L*. *Fifty Per Cent Proof: An Anthology of Mathematics Humour.* Keith Selkirk, William Wynne Willson. MAA, 1989, 87 pp, (P). [ISBN: 0-7487-0166-4] Chock-full of humorous stories, quotes, cartoons, misstatements, and jokes. Good evidence of something most of us (all of us?) already knew—mathematics can be very funny, especially when it isn't trying to. MPR

General, S, P, L. *Bias in Quantifying Judgments.* E.C. Poulton. Lawrence Erlbaum Associates, 1989, xxi + 304 pp, $49.95. [ISBN: 0-86377-105-X] Interesting but technical study of the various types of bias that can result when people make quantitative judgments. Includes both theory and practical illustrations, together with ways to avoid particular biases. Good set of references. RSK

General, S, L*.** *Paradoxes of Gambling Behaviour.* Willem Albert Wagenaar. Essays in Cognitive Psychology. Lawrence Erlbaum Associates, 1988, vii + 126 pp, $24.95. [ISBN: 0-86377-080-0] Psychological and statistical investigation of why people gamble, especially in the face of systematic losses. Based on extensive research with roulette, blackjack, and lottery players. One set of experiments clearly indicates that gamblers view chance and luck as different concepts (in roulette, e.g., chance determines the outcome of the wheel, but luck influences the player's choice of a number). Also includes a chapter on the legal problem of distinguishing between games of chance and games of skill. RSK

General, L.** *Envisioning Information.* Edward R. Tufte. Graphics Pr, 1990, 126 pp, $48. A magnificant sequel to Tufte's classic *The Visual Display of Quantitative Information* (TR, March 1984), this volume features maps, time tables, diagrams, charts—even computer screens—in order to illustrate wise principles of information design: "self-effacing displays intensely committed to rich data." A beautiful book, printed in six colors, whose own page layout is itself a pleasure to behold. LAS

General, S, L. *Figuring: The Joy of Numbers.* Shakuntala Devi. Penguin Books, 1990, 157 pp, £3.99 (P). [ISBN: 0-14-011850-0] Number curiosities and insider tricks for doing mental calculations with the four basic operations of addition, substraction, multiplication, and division. LCL

General, P*, L*. *Using Writing To Teach Mathematics.* Ed: Andrew Sterrett. MAA Notes No. 16. MAA, 1990, xvii + 139 pp, $15 (P). [ISBN: 0-88385-066-4] An important compendium of thirty-one diverse experiences in "writing to learn" mathematics—an overdue idea whose time has finally come. Authors report on very different approaches (from writing proofs to writing mathematical autobiographies; from keeping personal journals to preparing formal papers), operating at various curricular levels. "I need to explain this to myself," wrote one student, "so I can understand it." LAS

General, L. *Guidelines for the Use of Calculators in Competitions.* M. Kathleen Heid, *et al.* NCTM, 1989, v + 24 pp, $9.60 (P). [ISBN: 0-87353-286-4] Practical suggestions on such issues as deciding how and when calculators might be used, and on writing appropriate questions (including examples). Essential reading for all mathematics educators. LCL

Elementary, T(13). *Introductory Geometry: A Brief Course with Reasoning Skills.* Alan Wise. Harcourt Brace Jovanovich, 1990, viii + 132 pp, $6 (P). [ISBN: 0-15-546508-2] A guided workbook covering basic geometric concepts for college developmental mathematics courses. JNC

Precalculus, T*(13: 1). *Trigonometry.* Charles W. Gantner, Thomas E. Gantner. PWS-Kent, 1990, xii + 388 pp, $29.50. [ISBN: 0-534-92158-2] A review of geometrical concepts precedes coverage of traditional topics in right triangle and analytic trigonometry. Includes a nice historical introduction and interspersed historical notes, Basic programs for numerical computations and graphing, and comments about calculator usage. JNC

Finite Mathematics, T(13-14: 1), L. *Finite Mathematics.* Lawrence E. Spence, Charles Vanden Eynden, Daniel Gallin. Scott Foresman, 1990, xvi + 591 pp, $34. [ISBN: 0-673-38582-5] A straightfor-

ward presentation of the conventional material. Topics include linear programming, finance, probability and statistics, Markov chains, game theory. Good selection of examples and exercises. Index of applications. JS

Finite Mathematics, T(13: 2). *Applied Mathematics for the Management, Life, and Social Sciences.* Lawrence E. Spence, Charles Vanden Eynden, Daniel Gallin. Scott Foresman, 1990, xvi + 1049 pp, $40. [ISBN: 0-673-18837-X] Intuitive but honest. Stresses concepts over techniques; wide variety of applications. Algebra review throughout; exponential and logarithmic functions interspersed in several chapters to provide early introduction. Calculator access assumed, but not essential. Numerous exercises; practice problems with complete solutions at book's end. Appealing appearance. Supplements and aids are available. JDEK

Education, P. *From Gatekeeper to Gateway: Transforming Testing in America.* National Commission on Testing and Public Policy (Boston College, Chestnut Hill, MA), 1990, xiii + 59 pp, (P). An important and *very* timely polemic against the overuse, misuse, and abuse of multiple choice tests. Recommendations stress new objectives for testing—"to promote the development of all human talent" and to ensure greater public accountability. LAS

Education, P, L. *Soviet Studies in Mathematics Education.* Transl: Joan Teller. NCTM, 1990, $25 each (P), [ISBN: 0-87353-289-9] set. *Volume 1: Sets of Mathematics Teaching Aids,* M. Ya. Antonovskii, *et al.,* xix + 377 pp; *Volume 2: Types of Generalization in Instruction,* V.V. Davydov, xviii + 438 pp. First two volumes in a series of translations of Soviet literature on research in the psychology of mathematics instruction, especially on the dynamics of mental activity and principles of the learning process. *Volume 1,* originally published in 1971, provides a theoretical foundation for the use of instructional media in elementary mathematics lessons. *Volume 2,* originally published in 1972, conveys—in the context of a dialectical materialist theory of cognition—an innovative view of concept formation based on counterpoint between the "abstract" and the "concrete." LAS

Education, P*, L*. *Making the Grade in Mathematics: Elementary School Mathematics in the United States, Taiwan, and Japan.* Harold W. Stevenson, *et al.* NCTM, 1990, iv + 34 pp, $9.50 (P). [ISBN: 0-87353-292-9] A concise summary in plain language of the authors' comparative research from three countries (Taiwan, Japan, U.S.) of children's mathematics achievement, their attitudes about the role of work vs. ability, and their mothers' role and attitudes concerning mathematics education. Dispels several widespread myths (e.g., that Japanese and Chinese success in mathematics is due to rote learning, superior intelligence, or unrelenting parental pressure), and documents several key differences (e.g., educational priority of mathematics in school, support by parents at home). LAS

Education, P*, L.** *On the Shoulders of Giants: New Approaches to Numeracy.* Ed: Lynn Arthur Steen. National Academy Pr, 1990, vii + 232 pp, $17.95. [ISBN: 0-309-04234-8] A collection of five potential "strands" for school mathematics that illustrate how "mathematical ideas can be developed from informal childhood exploration through formal school and college study." Authors Thomas Banchoff, James Fey, David Moore, Marjorie Senechal, and Ian Stewart illustrate through five essays (on dimension, quantity, uncertainty, shape, and change) the potential of deep and powerful mathematics to stimulate rich and inviting school curricula.

Education, P. *Mathematics Contests: A Guide for Involving Students and Schools.* Frederick O. Flener. NCTM, 1990, iv + 115 pp, $9.95 (P). [ISBN: 0-87353-282-1] Some rationale for mathematics competitions, some hints for coaches, descriptions of the various forms that competitions take (timed individual or team events, relays, oral events, power questions ...), and a nice sampling of questions appropriate to these various forms. AWR

Education, P, L*. *A Sourcebook for College Mathematics Teaching.* Ed: Alan H. Schoenfeld. MAA, 1990, ii + 68 pp, $10 (P). [ISBN: 0-88385-068-0] A timely collection of pointers to improve undergraduate mathematics instruction: goals, curriculum guides, resources, advising, instructional techniques, use of computers, and evaluation. Individual chapters are filled with references to research findings and other sources of supplementary information. "Our goal is to inform. This is the year of national dialogue, and informed dialogue is the best kind." LAS

History, S, P*, L.** *Norbert Wiener, 1894-1964.* P.R. Masani. Birkhauser Verlag, 1990, 416 pp, $63. [ISBN: 0-8176-2246-2] A fascinating, thoroughly documented reflection on the intellectual and personal character of Norbert Wiener—a "universal thinker of colossal proportion." Biography; mathematics (Brownian motion, harmonic analysis, statistical physics, cybernetics); medicine (physiology, prosthetics, molecular biology); defense (WWI, anti-aircraft fire control, war games); religion (sin and evil, myth, catharsis, redemption); literature and art; and personal prejudices and blind spots. Includes extensive quotations from correspondence with many Wiener contemporaries (e.g., John von Neumann, T.S. Eliot, Walter Reuther), and many useful reference lists (vita, Ph.D. students, Wiener bibliography, references, index to over 500 names, and subject index). Presents Wiener's ideas in relief against the backdrop of his peers—other great thinkers both present and past. LAS

Logic, S(16-18), P, L. *An Introduction to Default Logic.* Philippe Besnard. Symbolic Computat. Springer-Verlag, 1989, xi + 208 pp, $39.50. [ISBN: 0-387-51566-6] Emphasis on formal development of default logic with a subsidiary theme of artificial intelligence applications. Begins with the need for default reasoning followed by an overview of formal logic as a knowledge representation scheme. Followed

by a summary of first order logic with nonmonotonic extensions to it. Chapters on default logic from the core of the text. Includes a chapter on circumscription, and a final chapter on alternate logic formalizations of nonmonotonic reasoning. References. RJA

Logic, P. *Lecture Notes in Mathematics-1388: Mathematical Logic and Applications.* Eds: J. Shinoda, T.A. Slaman, T. Tugué. Springer-Verlag, 1989, 222 pp, $20 (P). [ISBN: 0-387-51527-5] Proceedings of the 87th Meeting on Mathematical Logic and its Applications held in Kyoto, Japan during August 3-6, 1987. List of meeting participants. RJA

Logic, T(14-18: 1), S. *Systems of Logic.* Norman M. Martin. Cambridge University Pr, 1989, xi + 320 pp, $44.50; $15.95 (P). [ISBN: 0-521-36589-9; 0-521-36770-0] Assumes reader has had an introduction to logic. Studies and compares from a primarily syntactic point of view the characteristics of different logical systems: structural elements of a system, derivation and proof, interpretation and satisfaction, entailment and equivalence. Chapter exercises. Appendix of historical and bibliographical remarks. RJA

Logic, P. *Accessible Categories: The Foundations of Categorical Model Theory.* Michael Makkai, Robert Paré. Contemp. Math., V. 104. AMS, 1989, viii + 176 pp, $31 (P). [ISBN: 0-8218-5111-X] Introduction to categorical model theory defined by the authors as "the study of set-valued models of possibly infinitary first order theories by means of the conceptual tools of category theory." KES

Foundations, T(13-14: 1), L. *A Transition to Advanced Mathematics, Third Edition.* Douglas Smith, Maurice Eggen, Richard St. Andre. Brooks/Cole, 1990, xiii + 285 pp, $38. [ISBN: 0-534-12234-5] Extensive reworkings in Chapter 1 and 5; new optional section in Chapter 1 expands on examples of proof techniques; section on graphs in Chapter 3 rewritten (*First Edition,* TR, April 1984; *Second Edition,* TR, March 1987). LC

Foundations, T(16-18: 1), L. *Constructivism in Mathematics, An Introduction, Volume 1.* A.S. Troelstra, D. van Dalen. Stud. in Logic & Found. of Math., V. 121. North-Holland (US Distr: Elsevier Science), 1988, xx + 356 pp, $33 (P). [ISBN: 0-444-70506-6] First volume of two-volume text. Covers logical formalism for intuitionism, intuitionistic arithmetic, constructive treatment of real numbers, examples of constructive analysis. Assumes familiarity with classical predicate logic. KES

Graph Theory, T(12-14), S*, L.** *Graphs and Their Uses.* Oystein Ore. Revised by Robin J. Wilson. New Math. Lib., V. 34. MAA, 1990, viii + 153 pp, $14.50 (P). [ISBN: 0-88385-600-X] A classic reborn, with updated terminology, modern notation, and new material on interval graphs, travelling salesman problem, shortest route problems, and map-coloring. A brief contemporary bibliography points readers to sources of further information. A marvelous resource for math clubs, freshman seminars, and independent study. (Original edition is NML V. 10, TR, June-July 1976.) LAS

Combinatorics, T(16-17: 1), L. *Generating Functionology.* Herbert S. Wilf. Academic Pr, 1990, viii + 184 pp, $29.95. [ISBN: 0-12-751955-6] Introduction to generating functions and their applications in enumeration. For example, the generating function of a sequence can be used to find an exact formula, averages, recurrence relations, and prove facts about unimodality or convexity. In addition to the combinatorial aspects of generating functions, this text discusses asymptotic methods. LC

Combinatorics, P. *Tools and Fundamental Constructions of Combinatorial Mathematics.* Heinz Lüneburg. Bibliographisches Institut, 1989, 525 pp. [ISBN: 3-411-03194-8] The author's underlying belief is that induction and recursion are the most important tools of combinatorial mathematics. Thus he begins with Dedekind triples and studies them in depth, using them for instance to define finite sets. Topics covered include Gray codes, the symmetric group, the marriage theorem, independence structures, and partitions of finite sets. LC

Discrete Mathematics, T(13-14: 1, 2), L. *Discrete Mathematics with Applications.* Susanna S. Epp. Wadsworth, 1990, xx + 882 pp, $51.50. [ISBN: 0-534-09630-1] Similar in outline to most of the new texts, the treatment here is distinguished by a leisurely approach and attention to detail. There are two chapters on logic, two on methods of proof, and one on sets before counting is first encountered in Chapter 6 on page 315. Further chapters on functions, recursion, algorithms, and relations before graphs are introduced in Chapter 11, page 677. Copious examples and exercises with solutions and hints; index. JS

Number Theory, T*(18), S*, P.** *Class Field Theory.* Emil Artin, John Tate. Adv. Book Classics Ser. Addison-Wesley, 1990, xxxviii + 259 pp, $29.95. [ISBN: 0-201-51011-1] A reprint of the 1967 publication of Chapters 5-12 (global theory) of the notes from the Princeton seminar on Class Field Theory. Unfortunately it is a photoreproduction of the original typewritten text (with many handwritten symbols), but it is still a classic, typeset or not. "Special Preface" includes a list of eight related books published since the original printing. SB

Number Theory, P. *Automorphic Forms and Geometry of Arithmetic Varieties.* Eds: K. Hashimoto, Y. Namikawa. Adv. Stud. in Pure Math., V. 15. Academic Pr, 1989, xvi + 523 pp, $79.50. [ISBN: 0-12-330580-2] A collection of papers by Japanese researchers on dimension formulas for spaces of automorphic forms. The articles largely overlap a series of symposia held during 1985 and 1986. GG

Linear Algebra, T(14: 1). *Elementary Linear Algebra.* Harvey Gerber. Brooks/Cole, 1990, xiii + 498 pp, $42. [ISBN: 0-534-11574-8] A straightforward introductory text which includes historical notes, applications, and challenging supplementary exercises. Chapters cover systems of linear equations; matrices; determinants; R^2, R^3 and R^n; algebra of R^n; vec-

tor spaces; linear transformations; eigenvalues; linear programming. JNC

Linear Algebra, T*(14: 1). *Linear Algebra, Third Edition.* Michael O'Nan, Herbert Enderton. Harcourt Brace Jovanovich, 1990, ix + 461 pp, $35. [ISBN: 0-15-551008-8] Emphasis on geometrical point-of-view with gradual increase in level of abstraction and rigor. Computational algorithms feature Gaussian elimination and reduction to row-echelon form. In black and white, slowly paced, honest, and correct, from the concrete to the abstract. Few frills. Big plusses are the non-threatening writing style and the anticipation of customary hurdles. Numerous problems both routine and text-extending, with answers to many including evens. Solutions manual available. A textbook for those not in a hurry and without a need for excessive abstraction. (*First Edition,* TR, May 1971; Extended Review, June-July 1973; *Second Edition,* TR, June-July 1977.) JDEK

Linear Algebra, T(13-14). *Linear Algebra: An Introduction with Concurrent Examples.* A.G. Hamilton. Cambridge University Pr, 1989, x + 328 pp, $22.95 (P); $69.50. [ISBN: 0-521-31042-3; 0-521-32517-X] The first 40% of the book covers linear equations and vector geometry in two- and three-dimensions. The remainder is devoted to the concepts of linear algebra (basis, dimension, linear maps, eigen-theory, diagonalization, orthogonality, and quadratic forms) in R^n. An interesting feature is the placing of examples and exercises on the left-hand pages and the text on the right-hand pages. SG

Linear Algebra, T*(14-15: 1). *Linear Algebra, Second Edition.* John B. Fraleigh, Raymond A. Beauregard. Addison-Wesley, 1990, xiii + 571 pp, $35.16. [ISBN: 0-201-11949-8] Changes from *First Edition* (TR, April 1988) include expanded exercise sets (including ten-part True-False sections), examples from calculus, a chapter devoted to complex scalars, gradual presentation of linear transformations ($T(\mathbf{x}) = A\mathbf{x}$ in section 1.5, general transformations in 2.9), orthogonality chapter reorganized. A unique feature of this text is that eigenvalues, eigenvectors, and their applications are introduced in Chapter 4 (and recur in Chapters 5 and 8), not relegated to the last topic of the course. Although not intended to be overly rigorous, the text and exercises contain their fair share of proofs. SB

Calculus, T(13: 2). *Calculus for Management and the Life and Social Sciences, Second Edition.* Donald L. Stancl, Mildred L. Stancl. Richard D Irwin, 1990, xviii + 892 pp, $43.95. [ISBN: 0-256-08251-0] No specific revisions from *First Edition* (TR, December 1988) are cited. Includes chapters on probability and differential equations, in addition to standard calculus topics. No trigonometric functions appear until page 700. Chapters 2-12 end with supplements containing enrichment material. Supplements to the text include solution manuals, computer manual and disks, and test bank. SB

Calculus, T(13: 2-3). *Calculus with Analytic Geometry, Fourth Edition.* Robert Ellis, Denny Gulick. Harcourt Brace Jovanovich, 1990, xvii + 1125 pp, $44. [ISBN: 0-15-505687-5] A traditional approach with a few minor changes including the use of tangent lines and velocity to motivate limits, a new section on integration using tables, and a new chapter on complex numbers and functions. (*First Edition,* TR, August-September 1978; *Third Edition,* TR, October 1986; *Alternate Edition,* June-July 1988.) JNC

Calculus, T(13: 3). *Calculus: One Variable, Sixth Edition.* S.L. Salas, Einar Hille. Wiley, 1990, xiii + 777 pp, $46.95. [ISBN: 0-471-61626-5] Main difference between this and previous editions is the increased presence of applications scattered throughout the text. (*First Edition,* TR, June-July 1971; *Second Edition,* TR, October 1974; *Third Edition,* TR, June-July 1978 and December 1978; *Fourth Edition,* TR, June-July 1982; *Fifth Edition,* TR, November 1986.) KS

Calculus, T*(13: 2-3). *Calculus with Analytic Geometry, Third Edition.* John B. Fraleigh. Addison-Wesley, 1990, xv + 1032 pp, $45.56. [ISBN: 0-201-50363-8] This edition includes a graphical approach, historical notes, a streamlined treatment of limits, earlier treatment of trigonometric, exponential, and logarithmic functions, coverage of Taylor's theorem before infinite series, and calculator and computer exercises with references to *Exploring Calculus,* an IBM compatible software package. (*First Edition,* TR, August-September 1980; *Second Edition,* TR, January 1986.) JNC

Calculus, T(13: 1, 2), S. *Applied Calculus.* John C. Hegarty. Wiley, 1990, xv + 608 pp, $42.95. [ISBN: 0-471-61240-5] An informal, intuitive development of the rudiments of differential and integral calculus; applications, which are many and brief, lean heavily to economics. Logarithmic and exponential functions are introduced late; trigonometric functions not at all. Computation, not rigor, is stressed: e.g., the fundamental theorem appears, but not the mean value theorem. Computer and numerical methods are not emphasized. Most sections have large exercise sets, with many word problems. PZ

Real Analysis, T?(14). *Introduction to Real Analysis.* Robert L. Brabenec. PWS-Kent, 1990, xiv + 294 pp, $35. [ISBN: 0-534-92162-0] Essentially a standard first course text in real analysis. The author writes in a "user-friendly" style which perhaps makes the material more accessible to many students. However, the author is too casual in many ways, often to the point of error. For example: the first two instances of sequences are incorrectly used. MPR

Real Analysis, P. *The Ergodic Theory of Discrete Groups.* Peter J. Nicholls. London Math. Soc. Lect. Note Ser., V. 143. Cambridge University Pr, 1989, xi + 221 pp, $34.50 (P). [ISBN: 0-521-37674-2] A presentation of the classical work in this area. Also an introduction to the recent theory of measures on the limit set of a discrete group. Work of S.J. Patterson, D. Sullivan, and others. MLR

Complex Analysis, S(17-18), P. *Elements of the*

Theory of Elliptic Functions. N.I. Akhiezer. Transl. of Math. Mono., V. 79. AMS, 1990, vii + 237 pp, $93. [ISBN: 0-8218-4532-2] A translation of the 1970 Russian edition. This book, intended for engineers, presents the basics of the classical theory of elliptic functions: modular, Weierstrass, theta, and Jacobi functions; transformations of elliptic functions, and elliptic integrals. Includes tables of elliptic integrals. No exercises. SG

Complex Analysis, P. *Convolutional Calculus.* Ivan H. Dimovski. Math. & Its Applic., V. 43. Kluwer Academic, 1990, xi + 184 pp, $79. [ISBN: 0-7923-0623-6] Starting with Mikusiński's approach to Heaviside's operational calculus, the convolution is used as a means of unifying and extending these methods to first and second order differential operators. This is a book by an author who has been a key contributor to the new ideas here expounded. AWR

Complex Analysis, T(17-18:1). *Boundary Value Problems.* F.D. Gakhov. Transl: I.N. Sneddon. Dover, 1990, xix + 561 pp, $12.95 (P). [ISBN: 0-486-66275-6] Written on the basis of lectures given in 1947 so that "clarity and simplicity of style ...has taken preference over mathematical rigour," this book follows the spirit of historical development. A chapter on the Cauchy integral sets the stage for the Riemann Boundary Value Problem (Chapter II) and Singular Integral Equations (Chapter III), followed by the Hilbert Boundary Value Problem and Singular Integral Equations (Chapter IV). There is a final chapter on Generalized Boundary Value Problems. (1966 Addison-Wesley edition, TR, December 1967; Extended Review, November 1968.) AWR

Differential Equations, P, L. *The Asymptotic Solution of Linear Differential Systems: Applications of the Levinson Theorem.* M.S.P. Eastham. London Math. Soc. Mono. New Ser., V. 4. Clarendon Pr, 1989, x + 241 pp, $63. [ISBN: 0-19-853299-7] Levinson's 1948 theorem identified asymptotic solutions (for large x) of a broad class of differential equations of the form $y'(x) = f(x)y(x)$. This monograph shows how many of the asymptotic results that have been obtained since 1948 can be deduced directly from the Levinson theorem by certain well-defined transformation techniques. LAS

Partial Differential Equations, P. *Recent Topics in Nonlinear PDE IV.* Eds: Masayasu Mimura, Takaaki Nishida. Math. Stud., V. 160. North-Holland (US Distr: Elsevier Science), 1989, vii + 243 pp, $97.50 (P). [ISBN: 0-444-88087-9] Twelve papers from the fifth meeting on nonlinear partial differential equations held at Kyoto University, January 1988. MPR

Partial Differential Equations, P. *Scattering Theory, Revised Edition.* Peter D. Lax, Ralph S. Phillips. Pure & Appl. Math., V. 26. Academic Pr, 1989, xii + 309 pp, $44.95. [ISBN: 0-12-440051-5] Changes from the *First Edition* (TR, December 1967): Epilogue outlines new directions in the subject and results settled since the publication of the previous edition and includes a new list of references. MLR

Numerical Analysis, S(17-18), P, L. *Finite Difference Methods (Part 1), Solution of Equations in R^n (Part 1).* Eds: P.G. Ciarlet, J.L. Lions. Handbook of Numerical Analysis, V. 1. North-Holland (US Distr: Elsevier Science), 1990, vii + 652 pp, $107.25. [ISBN: 0-444-70366-7] The first volume of a series entitled "Handbook of Numerical Analysis" consisting of self-contained, accessible yet rigorous expository articles. This volume contains three articles covering finite difference methods for linear parabolic equations, splitting and alternating direction methods, and methods for solving least squares problems. SM

Operator Theory, T, P. *Positive Linear Systems: The Method of Positive Operators.* M.A. Krasnosel'skij, Je. A. Lifshits, A.V. Sobolev. Sigma Ser. in Appl. Math., V. 5. Heldermann Verlag, 1989, vii + 354 pp, DM 88 (P). [ISBN: 3-88538-405-1] Methods for studying systems and problems involving equations in spaces which are partially ordered by some cone. First chapter introduces basic notions and facts; the remaining three chapters deal with applications to spectral properties, iteration procedures, and other topics. Includes some exercises. LC

Functional Analysis, P. *Fourier Analysis on Groups.* Walter Rudin. Classics Library. Wiley, 1990, ix + 285 pp, $26.95. [ISBN: 0-471-52364-X] Paperback reproduction of Rudin's 1962 classic. MLR

Analysis, P. *Analysis and Partial Differential Equations.* Ed: Cora Sadosky. Lect. Notes in Pure & Appl. Math., V. 122. Marcel Dekker, 1990, xxvii + 742 pp, $115 (P). [ISBN: 0-8247-8302-6] Collection in honor of Mischa Cotlar's seventy-fifth birthday containing papers in harmonic and complex analysis, functional analysis and operator theory, and partial differential equations. Also contains complete bibliography of Cotlar's works, and two nice summaries of his contributions (pre- and post-1955). KS

Algebraic Geometry, S(18), P. *Lecture Notes in Mathematics-1400: Mixed Motives and Algebraic K-Theory.* Uwe Jannsen. Springer-Verlag, 1990, xiii + 246 pp, $26.50 (P). [ISBN: 0-387-52260-3] Part I develops mixed motives for absolute Hodge cycles; Part II discusses algebraic cycles, K-theory, and extension classes; while Part III goes further into K-theory and cohomology. Appendices, bibliography. JS

Algebraic Geometry, P. *Lecture Notes in Mathematics-1414: Algebraic Geometry and Complex Analysis.* Ed: E. Ramírez de Arellano. Springer-Verlag, 1989, 180 pp, $17.60 (P). [ISBN: 0-387-52175-5] Proceedings of an August 1987 workshop held in Pátzcuaro, Mexico. GG

Differential Geometry, P. *Lecture Notes in Mathematics-1416: Géométrie Symplectique et Mécanique.* Ed: C. Albert. Springer-Verlag, 1990, 289 pp, $26.50 (P). [ISBN: 0-387-52191-7] Seventeen papers from a conference organized by the Universities of Avignon, Lyon, Marseille, and Montpellier held in May 1988. MPR

Topology, T(18), S, P. *Topological Fields.* Seth Warner. Math. Stud., V. 157. North-Holland (US Distr: Elsevier Science), 1989, xiii + 563 pp, $110.50. [ISBN: 0-444-87429-1] Book contains a comprehensive treatment of the fundamentals of topological field theory, but also goes far enough so that the reader is prepared to tackle current literature. Includes ample exercises, historical notes on each topic covered, and a comprehensive bibliography. KS

Dynamical Systems, S(17-18), P*, L*. *Chaotic Evolution and Strange Attractors.* David Ruelle. Cambridge U Pr, 1989, xi + 96 pp, $12.95 (P); $39.50. [ISBN: 0-521-36830-8; 0-521-36272-5] Uses time series analysis to investigate chaotic behavior of deterministic dynamical systems. This book is a series of short chapters, making it an easy book to work through. No exercises. Intended for scientists in a wide range of fields, it assumes a relatively sophisticated understanding of differential equations. Provides a concise introduction to strange attractors which is suitable for those familiar with nonchaotic dynamical systems. SM

Dynamical Systems, T(17-18: 2). *Nonlinear Differential Equations and Dynamical Systems.* Ferdinand Verhulst. Universitext. Springer-Verlag, 1990, ix + 277 pp, $22 (P). [ISBN: 0-387-50628-4] Intended to bridge the gap between introductory graduate ordinary differential equations texts and research literature. Reasonably terse, but achieves its goal. Plenty of examples and exercises with solutions. JO

Probability, S, P*, L.** *The Empire of Chance: How Probability Changed Science and Everyday Life.* Gerd Gigerenzer, *et al.* Cambridge University Pr, 1989, xviii + 340 pp, $44.50. [ISBN: 0-521-33115-3] A narrative account of the reconciliation of *Fortuna,* the muse of unpredictability, with *Scientia,* the guiding hand of science. Three historical chapters are followed by three in-depth explorations of the evolution of probabilistic ideas in biology, physics, and psychology; two concluding chapters survey the role of numbers (hence chance) in the modern world (from baseball to medicine) and the implications of chance for inference, determinism, and other philosophical issues. Effective non-technical exposition with virtually no formulas. LAS

Stochastic Processes, T(17-18: 1, 2), S, P, L*. *Stochastic Processes.* J.L. Doob. Classics Library. Wiley, 1990, vii + 654 pp, $32.95 (P). [ISBN: 0-471-52369-0] Paperback reprint of 1953 original edition as part of Wiley's series of reprints of its own classic publications. This text, the first systematic survey of stochastic processes, permits "no compromise" with the mathematics of probability. Covers Markov processes, martingales, and stationary processes. A concluding supplement provides a concise survey of the necessary background in measure theory. LAS

Stochastic Processes, T(18: 1), P. *Stochastic Integration and Differential Equations, A New Approach.* Philip Protter. Applic. of Math., V. 21. Springer-Verlag, 1990, x + 302 pp, $49. [ISBN: 0-387-50996-8] An advanced text dealing with the integration of stochastic processes. Assumes a prior knowledge of stochastic processes through the study of martingales. The novelty of this work lies in its definition of a semimartingale as a "good" integrator of stochastic processes. SM

Stochastic Processes, T(17: 1), S, P. *The Elements of Stochastic Processes with Applications to the Natural Sciences.* Norman T.J. Bailey. Classics Library. Wiley, 1990, xi + 249 pp, $24.95 (P). [ISBN: 0-471-52368-2] Reprint of a text originally published in 1964, based on a set of lectures given by the author at John Hopkins Department of Biostatistics. Stochastic models are covered from the heuristic approach of applied mathematics, rather than through rigorous probability theory. Knowledge of probability, statistics, matrix algebra, complex variable theory, and differential equations is assumed. Covers generating functions, Markov chains, random walks, birth-and-death processes, queues, diffusion processes, and non-Markovian processes. Problems at the end of each chapter; not many applied problems. JJ

Elementary Statistics, T*(14-17: 1, 2), C. *Practical Business Statistics with StatPad.* Andrew F. Siegel. Richard D Irwin, 1990, xxx + 962 pp, $49.95. [ISBN: 0-256-07313-9] Designed primarily for MBA-level statistics courses. In addition to the usual topics, including multiple regression and nonparametrics, the book also has chapters on report writing, time series, and quality control. Contains many interesting real-world examples. Special feature is the inclusion of a disk containing StatPad, statistical software for use with the Lotus 1-2-3 spreadsheet, and exercises based on a data set which can be solved using StatPad or some other statistical package. A 138-page appendix contains the StatPad reference manual and tutorial. RSK

Elementary Statistics, T(14: 1, 2). *Introductory Statistics, Fifth Edition.* Thomas H. Wonnacott, Ronald J. Wonnacott. Wiley, 1990, xvi + 711 pp, $44.95. [ISBN: 0-471-61518-8] Revision of the 1985 *Fourth Edition (First Edition,* TR, October 1969; *Second Edition,* TR, August-September 1972; *Third Edition,* TR, March 1978). Includes new sections on graphs and common sense, expected value in bidding, the bootstrap, and robust estimation. Claimed extensive use of Minitab is minimal. Each chapter concludes with a "Final Challenge," a practical problem based on a real situation. RSK

Elementary Statistics, T(13-16: 1, 2). *Modern Data Analysis: A First Course in Applied Statistics.* Lawrence C. Hamilton. Brooks/Cole, 1990, xv + 684 pp, $43.25. [ISBN: 0-534-12846-7] Standard introductory text on applied statistics for nonmathematics majors. Does not require calculus. Emphasizes applications over theory, presenting many interesting data sets and graphical techniques. Covers typical topics: graphing data, summary statistics, introduction to probability, random sampling, the normal distribution (binomial and Poisson, also),

hypothesis testing and confidence intervals on means and proportions (but not variances), chi-square goodness of fit type tests, two-sample tests, nonparametrics, and regression (simple and multiple). Unfortunately, the text contains no computer instruction on how to use a statistical package for graphing and data analysis. There is a companion text called *Statistics with STATA* for this. The text only contains a short section on designing experiments and analysis of variance. JJ

Statistics, S*, L*. *Statistics in Psychology: An Historical Perspective.* Michael Cowles. Lawrence Erlbaum Associates, 1989, xvi + 218 pp, $36. [ISBN: 0-8058-0031-X] Survey of the historical development of classical statistical techniques found in basic texts on statistics for psychology, including interesting discussions of the personalities and controversies surrounding these developments. Should be of interest to anyone studying statistics. Good set of references. RSK

Statistics, P, L. *Handbook of Statistical Methods for Engineers and Scientists.* Ed: Harrison M. Wadsworth, Jr. McGraw-Hill, 1990, x + 706 pp, $76.50. [ISBN: 0-07-067674-7] Twenty sections by 25 authors beginning with basic statistics. Emphasis on quality control; also includes survey sampling, Bayesian analysis, experimental design, regression, simulation, robust and nonparametric methods, and time series. Good first reference for engineers. Includes section references. TH

Statistics, S, P. *Lecture Notes in Computer Science-420: Statistical and Scientific Database Management.* Ed: Z. Michalewicz. Springer-Verlag, 1990, v + 256 pp, $24.70 (P). [ISBN: 0-387-52342-1] Contains 16 research papers presented at the Fifth International Conference on Statistical and Scientific Databases held at the University of North Carolina, April 3-5, 1990. The papers address a wide range of issues related to storing and managing large collections of numerical data, especially statistical data related to the fields of geography, economics, and the physical sciences. GMS

Programming, T(14-18: 1), S. *An Introduction to Functional Programming Through Lambda Calculus.* Greg Michaelson. Intern. Comput. Sci. Ser. Addison-Wesley, 1989, xv + 320 pp, (P). [ISBN: 0-201-17812-5] Begins with the origins of functional programming and the differences between it and imperative programming. Investigates pure λ-calculus before proceeding to boolean values, operations, numbers, conditionals, and functions, along with representations of each. Lists, trees, evaluation orders, and termination make up the second third. Programming in standard ML and Lisp come last. Exercises and answers; bibliography. RJA

Programming, T(13-18), S. *Practical Prolog.* Richard J. Young. Van Nostrand Reinhold, 1989, xxi + 488 pp, $32.95 (P). [ISBN: 0-442-25934-4] A thorough, well-organized introduction to MProlog. Layout and explanations conducive to learning. No routine exercises. Three appendices of predicates. Companion software available. RJA

Programming, T(13). *BASIC: An Introduction to Computer Programming, Fourth Edition.* Robert J. Bent, George C. Sethares. Ser. in Comput. Sci. Brooks/Cole, 1990, xiv + 384 pp, $35.25 (P). [ISBN: 0-534-12642-1] Revisions include expanded introductory chapter on problem solving, a new section on menu-driven programs and table processing, increased emphasis on the application of string variables, improved examples and exercises (*First Edition,* TR, March 1979). LC

Programming, T(15-18: 1), S. *LISP, Lore, and Logic: An Algebraic View of LISP Programming, Foundations, and Applications.* W. Richard Stark. Springer-Verlag, 1990, xviii + 278 pp, $38 (P). [ISBN: 0-387-97072-X] A combination of both the practical and the theoretical. Contains historical notes throughout. Syntax, programming style, and semantics of computation are followed by the algebra of interpreters, the lambda calculus, and the Lisp approach to artificial intelligence. Chapter exercises; answers to selected questions; bibliography; index. RJA

Computer Systems, P, L. *A Gentle Introduction to TEX: A Manual for Self-study.* Michael Doob. TEXniques, No. 12. TEXUsers Group (POB 9506, Providence, RI 02940), 1990, iv + 88 pp, $15 (P). As advertised—a concise, careful introduction for beginners, far less intimidating (and far less complete) than Knuth's *TEXbook,* to which page references occur throughout the margins of this primer. Characters, shapes, mathematics, alignment, errors, boxes—just enough to do simple things. If you are just beginning to use TEX, read this first. LAS

Theory of Computation, S, P. *First-Order Logic and Automated Theorem Proving.* Melvin Fitting. Texts & Monographs in Computer Science. Springer-Verlag, 1990, xv + 242 pp, $35. [ISBN: 0-387-97233-1] Text for researchers in computer science who are interested in the applications of formal logic to the topic of automated theorem proving. It covers work on proof methods, unification, and resolution so important to the area of logic-based programming languages. It also discusses other topics including equality, semantic tableaux, and heuristics. GMS

Artificial Intelligence, T(16), S. *Object-Oriented Simulation with Hierarchical, Modular Models: Intelligent Agents and Endomorphic Systems.* Bernard P. Zeigler. Academic Pr, 1990, xvii + 395 pp, $49.95. [ISBN: 0-12-778452-7] Describes a simulation system called DEVS (Discrete Event Simulation System). The system, written in a language called Scheme, is intended for use in simulating and studying intelligent systems in a realistic manner. It describes the system as well as the object-oriented model upon which it is based. It includes numerous examples and case studies. GMS

Artificial Intelligence, P. *The Acquisition of Strategic Knowledge.* Thomas R. Gruber. Perspect. in Artif. Intellig., V. 4. Academic Pr, 1989, xxiv + 311 pp, $39.95. [ISBN: 0-12-304754-4] Describes

a computer-assisted approach to obtaining human knowledge and formulating it for use in knowledge-based systems. The ASK program is described and details of its use presented. Includes techniques for knowledge elicitation, machine learning, knowledge engineering. Contains a chapter on evaluation of this second-generation artificial intelligence research. Bibliography. RJA

Artificial Intelligence, P. *Exemplar-Based Knowledge Acquisition: A Unified Approach to Concept Representation, Classification, and Learning.* Ray Bareiss. Perspect. in Artific. Intellig., V. 2. Academic Pr, 1989, xii + 169 pp, $24.95. [ISBN: 0-12-078260-X] Describes the Protos knowledge acquisition system, a learning apprentice that forms categories of common experiences and classifies new experiences into them. Overview of basic issues and approaches to them. Design and evaluation of Protos. Ends with related research and future work. Appendices; glossaries; bibliography. RJA

Computer Science, S(16-18), P. *Predicate Calculus and Program Semantics.* Edsger W. Dijkstra, Carel S. Scholten. Texts & Mono. in Comput. Sci. Springer-Verlag, 1990, xi + 220 pp, $39. [ISBN: 0-387-96957-8] A predicate transformer is a function from boolean structures to boolean structures. Main topics include the calculus of boolean structures, properties of predicate transformers, semantics of repetitions, and use of predicate transformers to define programming language semantics. RJA

Computer Science, T(14-18: 1, 2), L. *The Logical Basis for Computer Programming.* Zohar Manna, Richard Waldinger. Addison-Wesley. *Volume 1: Deductive Reasoning,* 1985, xix + 618 pp, $25.56 [ISBN: 0-201-18260-2]; *Volume 2: Deductive Systems,* 1990, xiii + 642 pp, $27.16. [ISBN: 0-201-18261-0] *Volume 1* begins with an introduction to mathematical logic, both propositional and predicate; treats theories with induction and includes axiomatic presentations of the main structures in computer science (strings, lists, trees, sets, etc.). *Volume 2* is more formal in its development and treats four main topics: well-founded induction, propositional logic, unification, and predicate logic. A deductive-tableau system is given for proving the validity of sentences. Chapter problems (solutions can be ordered). List of related textbooks and selected references. Symbol and general index. RJA

Computer Science, P. *Object-Oriented Concepts, Databases, and Applications.* Eds: Won Kim, Frederick H. Lochovsky. Addison-Wesley, 1989, xv + 602 pp. [ISBN: 0-201-14410-7] Divided into five areas: object-oriented concepts embedded in object-oriented programming languages and semantic data models; applications; database systems; architectural issues; directions for future research and development. Index. RJA

Applications, T(16-17: 2). *Mathematical Methods for Scientists and Engineers: Linear and Nonlinear Systems.* Peter B. Kahn. Wiley, 1990, xix + 469 pp, $54.95. [ISBN: 0-471-62305-9] Emphasizes perturbation techniques and asymptotic methods. Nonlinear problems and associated models in diverse fields. Long time in production. Classroom tested; informal and readable, in repetitive style, with the same examples recurring throughout. Prerequisites include calculus through differential equations, some mechanics, some linear algebra, but no complex variables. Develops tools of analysis for treatment of problems without closed-form solutions. For advanced undergraduates or beginning graduate students in physics, chemistry, engineering, and biology. Numerical and graphical solutions. Software packages from Lotus 1-2-3 and Math CAD. Modest number of exercises, with answers where appropriate, throughout the text. JDEK

Applications (Economics), T(15-17: 1), L. *Differential Equations, Stability and Chaos in Dynamic Economics.* W.A. Brock, A.G. Malliaris. Adv. Textbooks in Econ., V. 27. North-Holland (US Distr: Elsevier Science), 1989, xvi + 389 pp, $49. [ISBN: 0-444-70500-7] A text designed to take advanced students of economics through an elementary study of differential equations. The first half covers the usual topics: existence, uniqueness, dependence on initial conditions, linear systems, linearization and Lyapunov analysis of critical points. The second half covers applications in several areas of economic theory, including micro- and macro-economic policies and investment theory. The last chapter is an introduction to chaos in economic systems. Extensive bibliography. Exercises at the end of each chapter. SM

Applications (Physics), S(13-16), P*, L**.** *The New Ambidextrous Universe: Symmetry and Asymmetry from Mirror Reflections to Superstrings, Third Revised Edition.* Martin Gardner. WH Freeman, 1990, xiv + 392 pp, $19.95. [ISBN: 0-7167-2092-2] Third revision of Gardner's classic 1964 nontechnical exposition of the fall of parity. The 1979 *Second Edition* (TR, May 1979) added five chapters on the violation of time-invariance symmetry. This edition, in addition to minor changes throughout, adds five more chapters on twistors, monopoles, spin, and superstrings. Absorbing; enlightening; lucid; witty; inventive. An exemplar of science writing at its very best. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; SB: Steve Benson, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; TH: Timothy Hesterberg, St. Olaf; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; JO: Jeff Ondich, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, St. Olaf; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; KES: Kay E. Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; PZ: Paul Zorn, St. Olaf.

Addison-Wesley gives your students the best of math

Differential Equations: A Modeling Approach
Frank R. Giordano and Maurice D. Weir, *U.S. Military Academy and U.S. Naval Postgraduate School, respectively*

Now your students can clearly see how differential equations apply to realistic problems. This new text by Giordano and Weir focuses on mathematical modeling, showing students a natural progression of related ideas, and offering motivation to take the next step along the way. Unlike other texts that use a "solutions technique" approach, **Differential Equations: A Modeling Approach** puts into perspective how the various solutions techniques students learn actually fit together — the "why" as well as the "how-to."

Discrete Algorithmic Mathematics
Stephen B. Maurer and Anthony Ralston, *Swarthmore College and SUNY-Buffalo, respectively*

Turn to Maurer and Ralston for discrete mathematics with a basic difference: a unique algorithmic approach. The authors present discrete mathematics as a unified topic, by viewing it through an algorithmic "prism." A wide range of applications, flexible coverage of core discrete math topics, and extensive coverage of linear algebra also add to this text's substantial value.

Addison-Wesley
1 Jacob Way • Reading, MA • 01867
(617) 944-3700

From the publisher that's serious about math

▼
ALSO NEW FOR '91!

Developmental Mathematics

Basic Mathematics, Sixth Edition
Introductory Algebra, Sixth Edition
Intermediate Algebra, Sixth Edition
Mervin L. Keedy and Marvin L. Bittinger, *Purdue University and Indiana University-Purdue University at Indianapolis, respectively*

Concepts and Applications of Intermediate Algebra, Second Edition
Marvin L. Bittinger, Mervin L. Keedy, and David Ellenbogen, *Indiana University-Purdue University at Indianapolis, Purdue University, and St. Michael's College, respectively*

Elementary Algebra
Intermediate Algebra
Mark Dugopolski, *Southeastern Louisiana University*

Statistics

Introductory Statistics, Third Edition
Neil A. Weiss and Matthew Hassett, *both of Arizona State University*

Statistics Methods for Business and Economics, Fourth Edition
Donald L. Harnett and Ashok Soni, *both of Indiana University*

▼
OTHER TITLES OF INTEREST

Precalculus: Functions and Graphs
A. Robert Marshall, *Cuesta College*

Mathematics for Elementary School Teachers: A Problem Solving Approach
Joseph Newmark, *CUNY College of Staten Island*

Calculus: A Complete Course
Robert A. Adams, *University of British Columbia*

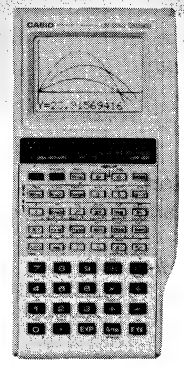
The Power to Project Your Ideas



The OH-7000G lets the instructor lead the class via a standard overhead projector.

Projectable Graphic Calculator.

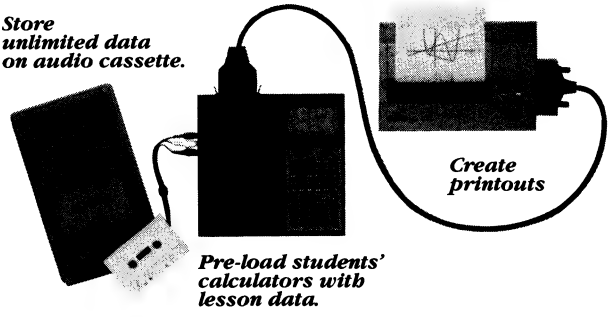
Casio has taken one of its most powerful graphic calculators and turned it into the most powerful **teaching tool** of its kind. The OH-7000G features a transparent, projectable screen so that teachers can lead their class just by using a standard overhead projector.



OH-7000G

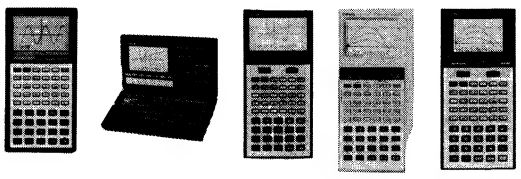
Since inventing the graphic calculator in 1985, Casio has created an entire family of exciting classroom products. Five different graphic calculator models are available, each offering **hundreds of powerful, built-in functions**.

With our **printers**, you can create "hard copy" of graphs and calculations for photocopying onto test papers or transparencies. The **FA-80 Data Interface** lets you store unlimited data on audio cassettes.



The possibilities are endless. That's why the **NCTM curriculum** calls for the use of graphic calculators. Many **new textbooks** are being published especially for use with this technology.

So, take a bold, new step and enter Casio's "Electronic Classroom" – a place where the solution is always clear.



CASIO

Creators of th Graphic Calculato

570 Mt. Pleasant Avenue, Dover, New Jersey 07

For more information about Casio's family of graphic calculators, and to place an order, call one of the educational distributors listed below.

ABM
800-234-3396
(In TX 214-869-0721)
(In MO 816-861-1113)
(In CA 213-946-6680)

Allied Communications
800-843-3660
(In PA 800-843-3660)

Allied National
800-999-8099
(In MI 313-543-1232)

Calculators Inc.
800-533-9921
(In MN 800-533-9921)

Carolina Wholesale
800-521-4600
(In NC 704-598-8101)

Colborn School Supply
800-548-7031
(In CO 303-778-1220)
(In MT 406-245-3158)

Copco
800-446-7021
(In OH 800-589-3006)

Dale Seymour Publications
800-872-1100
(In CA 800-222-0766)

Douglas Stewart
800-356-8370
(In WI 608-221-1155)

E.A.I.
800-272-0272
(In NJ 201-891-9466)

El Dorado Trading
800-227-8292
(In CA 415-494-1995)

Entertainment Marketing
800-347-2001
(In TX 713-995-4433)

Global Products
800-633-0633
(In IL 708-397-4944)

Longino Distributors
800-633-6224
(In NC 704-873-3282)

NASCO
800-558-9595
(In WI 414-563-2446)

Nemesis Distributing
800-940-7407
(In FL 305-477-8822)

Servco Pacific
(In HI 808-841-7566)

Tam's
800-421-5188
(In CA 800-244-5624)

Taylor Electric
800-558-6970
(In WI 800-242-8940)

Techline
800-777-3635
(In VA 703-389-0857)

Valley Business Machines
(In UT 801-969-6303)

Vistatech
800-847-9851
(In NY 212-254-9851)
(In CA 213-602-0277)

CASIO®

When was the last time a computer program helped you think about mathematics?

Call or write for our latest software catalog.
For PC or Macintosh.

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

New From Martin Gardner



MATHEMATICAL MAGIC SHOW

Martin Gardner

Martin Gardner published his first book in 1935. Since then he has charmed, puzzled, and delighted countless readers in his more than 409 books, among them a novel. Here is what reviewers have said about this book.

Highly recommended, but be warned—mathematical games can be addictive.

David Jones in *New Scientist*

Mathematical Magic Show begins with a chapter on nothing, and finishes with a chapter on everything. In between we visit most of the prime sites of recreational mathematics—game theory, factorial...puzzles, playing cards, finger arithmetic, Möbias bands, polyominoes, perfect numbers, the knight's tour, trees, and dice. Gardner always has new facts and ideas to add interest to even the most well-trodden areas. For example, he extends his discussion of the knight's tour to bring in the cook's tour (a cook travels three squares forward and then one square right or left) and then goes on to include camels, asps, and giraffes. The chapter on dice has some very useful hints on cheating at craps, and how not to get caught.

Harvey Mellar in *Times Literary Supplement*

Over the 30 or so years that (Martin Gardner's) column ran in Scientific American, he built up an enthusiastic readership which included practically every professional mathematician that I have met, and no doubt the countless many I have not. There must be a great many professionals whose own interest in the subject was awakened by his rare gift of being able to put across to the outsider the deepest developments in an extremely difficult and abstract subject, and to convey the enjoyment and excitement that lies within mathematics proper.

Keith Devlin in *The Guardian*

312 pp., Paper, 1990, ISBN 0-88385-445-1

List: $17.50 MAA Member: $14.50

Catalog Number MAGIC

ORDER FROM

The Mathematical Association of America

1529 Eighteenth Street, N.W.
Washington, D.C. 20036

NEW IN PAPERBACK FROM PRINCETON

Georg Cantor

His Mathematics and Philosophy of the Infinite

Joseph Warren Dauben

One of the greatest revolutions in mathematics occurred when Georg Cantor (1845–1918) promulgated his theory of transfinite sets. This revolution is the subject of Joseph Dauben's important study—the most thorough yet written—of the philosopher and mathematician who was once called a "corrupter of youth" for an innovation that is now a vital component of elementary school curricula.

"Historians of mathematics can only be grateful for the effort Professor Dauben has expended to create the synthesis of Cantor scholarship found in this book. But the book can, and I hope will, be read with profit by a far more extensive audience."
—*Thomas Hawkins, Historia Mathematica*

Now in paper: $14.95 ISBN 0-691-02447-2

Mathematics and Plausible Reasoning

Volume I: Induction and Analogy in Mathematics

Volume II: Patterns of Plausible Inference

G. Polya

Here the author of *How to Solve It* explains how to become a "good guesser." Marked by G. Polya's simple, energetic prose and use of clever examples from a wide range of human activities, this two-volume work explores techniques of guessing, inductive reasoning, reasoning by analogy, and the role they play in the most rigorous of deductive disciplines.

Volume I: $12.95 Paper ISBN 0-691-02509-6
Volume II: $12.95 Paper ISBN 0-691-02510-X

AT YOUR BOOKSTORE OR

Princeton University Press

41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900
ORDERS: 800-PRS-ISBN (777-4726)

PLANT THE SEEDS TO HIGHER LEARNING . . .

Being first is almost second nature at McGraw-Hill. Whether your students are enrolled in developmental math, precalculus, or calculus, McGraw-Hill's texts will inspire your students to learn.

NEW FOR 1991

The Streeter Series

BASIC MATHEMATICAL SKILLS WITH GEOMETRY
Streeter/Hutchison/Hoelzle

BEGINNING ALGEBRA, Form B
Streeter/Hutchison/Hoelzle

INTERMEDIATE ALGEBRA, Form B
Streeter/Hutchison/Hoelzle

COLLEGE ALGEBRA WITH TRIGONOMETRY
Rees/Sparks/Rees

PRECALCULUS: A Bridge to Calculus
Runyan/Runyan

EXPLORATIONS IN CALCULUS WITH A COMPUTER ALGEBRA SYSTEM
Small/Hosack

DISCOVERING CALCULUS WITH THE HP-28s AND THE HP-48sx
Smith/Minton

DISCRETE MATH AND ITS APPLICATIONS, 2/e
Rosen

DIFFERENTIAL EQUATIONS WITH APPLICATIONS, 2/e
Simmons

INTRODUCTION TO PARTIAL DIFFERENTIAL EQUATIONS, 2/e
Pinsky

FUNCTIONAL ANALYSIS, 2/e
Rudin

MATRIX THEORY
Goldberg

BRIDGE TO ABSTRACT MATHEMATICS, 2/e
Morash

For more information, please contact your local McGraw-Hill representative or write: McGraw-Hill College Division, Comp Processing and Control, P.O. Box 448, Hightstown, NJ 08520.

BACK-TO-BASICS INSTRUCTIONAL BOOKS

PERSONAL MATHEMATICS AND COMPUTING

Tools for the Liberal Arts

Frank Wattenberg

This book is a self-contained tutorial that can assist any reader with a background in high school mathematics in learning how to apply personal computing to enhance his or her understanding of modern quantitative methods. *The New Liberal Arts*

492 pp. $29.95 (December)

Original in Paperback

INVESTIGATIONS IN ALGEBRA

Albert A. Cuoco

"This erudite and witty book will be an invaluable resource for high-school math teachers and their students. It provides an in-depth opportunity for *doing* mathematics, not just learning *about* it."— *Laurence J. Davidson, Educator, author of Apple Logo Reference Manual*

350 pp. $45.00 cloth, $29.95 paper

APPROACHING PRECALCULUS MATHEMATICS DISCRETELY

Explorations in a Computer Environment

Philip G. Lewis

Lewis demonstrates how concepts of discrete mathematics can be introduced through the computer using the Logo™ language, making them easier to teach and more fun to learn.

557 pp. $35.00 cloth, $19.95 paper

To order IBM Logo™ call 1 800 IBM 2468
To order LCSI Logo II™ for Apple call Logo Computer Systems Inc.

To order books call toll free 1 800 356 0343

55 Hayward Street **THE MIT PRESS** **Cambridge, MA 02142**

U S Postal Service

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

Required by 39 U.S.C. 3685)

1A Title of Publication	1B PUBLICATION NO	2 Date of Filing
THE AMERICAN MATHEMATICAL MONTHLY	0 0 0 2 9 8 9 0	Sept. 20, 1990

3 Frequency of Issue	3A No. of Issues Published Annually	3B Annual Subscription Price
monthly except bi-monthly June/July and Aug/Sept	Ten	Library $120 Indiv. Member $30

4 Complete Mailing Address of Known Office of Publication *(Street, City, County, State and ZIP+4 Code) (Not printers)*

1529 Eighteenth St., N.W., Washington, D.C. 20036

5 Complete Mailing Address of the Headquarters of General Business Offices of the Publisher *(Not printer)*

1529 Eighteenth St., N.W., Washington, D.C. 20036

6 Full Names and Complete Mailing Address of Publisher, Editor, and Managing Editor *(This item MUST NOT be blank)*

Publisher *(Name and Complete Mailing Address)*

The Mathematical Association of America, 1529 Eighteenth St.,N.W., Washington, D.C. 20036

Editor *(Name and Complete Mailing Address)*

Prof. Herbert S. Wilf, Univ. of Pennsylvania, Dept. of Mathematics, Philadelphia, PA 19104

Managing Editor *(Name and Complete Mailing Address)*

Harry Waldman, MAA, 1529 Eighteenth St., N.W., Washington, D.C. 20036

7 Owner *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)*

Full Name	Complete Mailing Address
The Mathematical Association of America, Inc.	1529 Eighteenth St., N.W. Washington, D.C. 20036

8 Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities *(If there are none, so state)*

Full Name	Complete Mailing Address
none	

9 For Completion by Nonprofit Organizations Authorized To Mail at Special Rates *(DMM Section 423.12 only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [X] Has Not Changed During Preceding 12 Months
(2) [] Has Changed During Preceding 12 Months
(If changed, publisher must submit explanation of change with this statement.)

10 Extent and Nature of Circulation *(See instructions on reverse side)*	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A Total No. Copies *(Net Press Run)*	23,642	22,437
B Paid and/or Requested Circulation 1. Sales through dealers and carriers, street vendors and counter sales	0	0
2. Mail Subscription *(Paid and/or requested)*	19,902	18,873
C Total Paid and/or Requested Circulation *(Sum of 10B1 and 10B2)*	19,902	18,873
D Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies	898	830
E Total Distribution *(Sum of C and D)*	20,800	19,703
F Copies Not Distributed 1. Office use, left over, unaccounted, spoiled after printing	2,842	2,734
2 Return from News Agents	0	0
G TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	23,642	22,437

11. I certify that the statements made by me above are correct and complete

Signature and Title of Editor, Publisher, Business Manager, or Owner
Kay Hamond Membership Manager

PS Form 3526, Feb 1989 *(See reverse)*

Studies in Numerical Analysis,

Gene H. Golub, Editor
MAA Studies in Mathematics #24

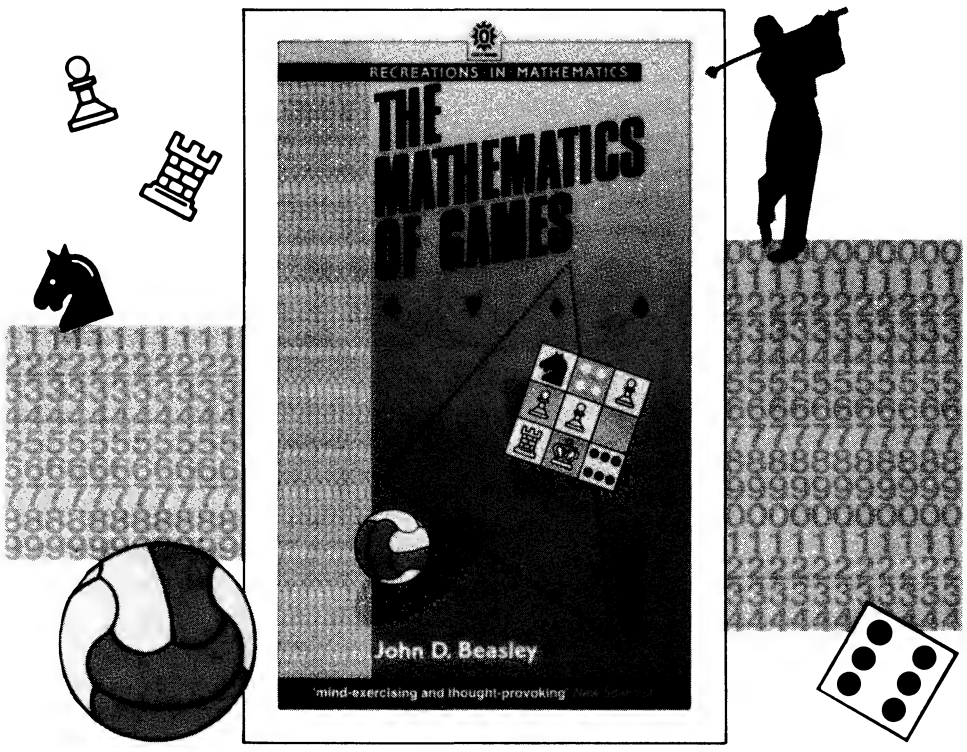
415 pp. Cloth, 1985,
ISBN-0-88385-126-1
List: $46.50
MAA Member: $34.50

This volume is a collection of papers describing the wide range of research activity in numerical analysis. The articles describe solutions to a variety of problems using many different kinds of computational tools. Some of the computations require nothing more than a hand-held calculator; others require the most modern computer. While the papers do not cover all of the problems that arise in numerical analysis, they do offer an enticing and informative sampling.

Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

NEW IN PAPERBACK!

"An entertaining excursion for the general reader with an interest in computational recreational mathematics, for teachers looking for examples, and for mathematics students looking for ideas to explore." ***— Choice***



From the luck of the die to the cerebral problems of chess, this entertaining and wide-ranging guide shows how simple mathematical analysis can throw unexpected light on games of every type. *The Mathematics of Games* outlines the general principles behind such classic puzzles as peg solitaire and Rubik's cube, explores the role of luck in ball games like golf and soccer, and examines how the study of even the most simple games can reveal some of the fundamental paradoxes of mathematics. 192 pp.; 94 illus. paper $7.95

At better bookstores or directly from

OXFORD PAPERBACKS

Oxford University Press • 200 Madison Avenue • New York • New York • 10016

New for 1991

PREALGEBRA
JOSHUA BERENBOM and UMESH NAGARKATTE
Available December 1990.

FUNDAMENTALS OF MATHEMATICS
MICHAEL SENTLOWITZ and PATRICIA L. HIRSCHY
Available December 1990.

FINITE MATHEMATICS
WILLIAM OWEN and FRED CUTLIP
Available in January.

ALGEBRA FOR COLLEGE STUDENTS
Third Edition
BERNARD KOLMAN and ARNOLD SHAPIRO
Just Published!

ABSTRACT ALGEBRA
DENNIS KLETZING
Available in January.

ELEMENTARY LINEAR ALGEBRA WITH APPLICATIONS
Second Edition
RICHARD O. HILL
Available December 1990.

from HBJ!

INTRODUCTORY ALGEBRA
Just Published!

INTERMEDIATE ALGEBRA
Both by RONALD HATTON and GENE SELLERS
Available January 1991.

MASTERING MATHEMATICAL CONCEPTS
Second Edition

ELEMENTARY ALGEBRA
Second Edition
Hardcover, 1991

INTERMEDIATE ALGEBRA
Second Edition
All by YOSHIKO YAMATO and MARY JANE CORDON
Hardcover, 1991
All Just Published!

CALCULUS One and Several Variables
ROBERT ELLIS and DENNY GULICK
Just Published!

SINGLE-VARIABLE CALCULUS WITH DISCRETE MATHEMATICS
JOHN A. FEROE and CHARLES STEINHORN
Available in January.

Also Just Published from HBJ...
Preliminary Editions of The Clemson Calculator Project's Enhancements for
- Single-Variable Calculus
- Multivariable Calculus
- Linear Algebra
- Differential Equations
- Introductory Statistics

HBJ
HARCOURT BRACE JOVANOVICH, INC.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648
(708) 647-8822

INDEX TO VOLUME 97, 1990
THE AMERICAN MATHEMATICAL MONTHLY

Title Index	949	Reviews by Author	955
Author Index	951	Index of Special Topics	956
Subject Index	953	Solutions	956
Reviews by Title	955	Problems Proposed	957
		Problems Solved	958

TITLE INDEX

Adding up to Powers, Edward A. Bender, Fred Kochman, and Douglas B. West, 139

An Alternative Proof of the Lindemann-Weierstrass Theorem, Frits Beukers, Jean-Paul Bézivin and Philippe Robba, 193

Ambiguous Numbers are Dense, Marko Petkovšek, 408

Another Proof of the Quadratic Reciprocity Theorem?, Richard G. Swan, 138

A Birthday Present for Ramanujan, M. D. Hirschhorn, 398

The Bitangent Sphere Problem, Peter J. Giblin and Donal B. O'Shea, 5

Cauchy-Bunyakowski-Schwarz Inequality Revisited, François Dubeau, 419

A Cayley Theorem for Boolean Algebras, Stephen L. Bloom, Zoltán Ésik and Ernest G. Manes, 831

A Condition for Flatness of Curves in R^n, Yaakov S. Kupitz and Micha A. Perles, 401

Continuous Functions and Connected Graphs, C. E. Burgess, 337

Convexity, Marcel Berger, 650

Conway's Tiling Groups, William P. Thurston, 757

A Counterexample on Vector Norms and the Subordinate Matrix Norms, Pieter P. N. De Groen, 406

A Counterexample in Galois Theory, Thomas W. Hungerford, 54

Curvature in the Eighties, Robert Osserman, 731

The Dark Side of the Möbius Strip, Gideon Schwarz, 890

A Decomposition Theorem for Finite Automata, Teresa L. Santa Coloma and Ralph P. Tucci, 51

Decomposition Theory in the Teaching of Elementary Linear Algebra, R. R. London and H. P. Rogosinski, 478

Diagonalization over Commutative Rings, R. Bruce Richter and William P. Wardlaw, 223

The Diophantine Equation $x^2 + y^m = z^{2n}$, Addendum, David W. Boyd, 411

The Dirichlet Problem for a Disk, David Minda, 220

Elementary Evaluations of $\int_0^\infty e^{-x^2}\,dx$, $\int_0^\infty \cos x^2\,dx$, and $\int_0^\infty \sin x^2\,dx$, Robert Weinstock, 39

The Ergodic Character of Sequences of Pedal Triangles, Peter D. Lax, 377

The Euclidean Algorithm Strikes Again, Stan Wagon, 125

Extending the Principal Axis Theorem to Fields Other Than $\mathbf{R}$, Stephen H. Friedberg, 147

A New Extension of the Derivative, Daniel Kopel and Michael Schramm, 230

Fibonacci Numbers and Aitken Sequences Revisited, M. J. Jamieson, 829

Finding the Finite Groups of Symmetries of the Sphere, Marjorie Senechal, 329

For Your Eyes Only, Gary A. Martin, 418

Functional Equations and L'Hôpital's Rule in an Exact Poisson Derivation, J. Aczél, 423

Game of Cards, Dynamical Systems, and a Characterization of the Floor and Ceiling Functions, P. Eisele and K. P. Hadeler, 466

Gear Trains and Continued Fractions, Jerold Mathews, 505

A Geometric Approach to l'Hôpital's Rule, Gianluca Gorni, 518

A Geometric Regularity From Order-Restricted Statistical Inference, Carolyn Pillers and Tim Robertson, 496

The Geometry of Harmonic Oscillators, Kenneth R. Meyer, 457

What Is Geometry?, Shiing-Shen Chern, 679

Greed Is (Sometimes) Not Enough, Anatole Beck, 289

A Guessing Game of Bill Sands, and Bernardo Recamán's Barranca, Richard K. Guy, 314

On Inequivalent Group Extensions, F. B. Cannonito, 317

An Infinite Product Expansion for the Square Root Function, Eric Wingler, 836

Integral Representation of a Finite Spike, R. E. Mirollo and S. H. Strogatz, 901

Inverse Functions and their Derivatives, Ernst Snapper, 144

On the Irrationality of π^2, D. Desbrow, 903

Set Isometries and Their Extensions in the Max Metric, Regina B. Cohen, 795

Isoptic Characterization of a Circle (Proof of a Conjecture of M. S. Klamkin), Johannes C. C. Nitsche, 45

The L^4 Norm of a Polynomial with Coefficients ± 1, J. S. Byrnes and Donald J. Newman, 42

The l_2 Matrix Norm and the Spectral Radius, Robert A. Leslie, 339

The Lebesgue Integral Without Measure Theory, A. Van Daele, 912

A Link Between the Jordan Curve Theorem and the Kuratowski Planarity Criterion, Carsten Thomassen, 216

Lucas's Theorem and Some Related Results for Extended Pascal Triangles, Richard Bollinger and Charles Burchard, 198

Magic Squares and Linear Algebra, Arno Van Den Essen, 60

Minimal Surfaces Based on the Catenoid, David Hoffman and William H. Meeks, III, 702

Mixing Property of the Pedal Mapping, Peter Ungar, 898

More on Mortality, Melven Krom and Myren Krom, 37

Musings on the Prime Divisors of Arithmetic Sequences, Patrick Morton, 323

A New Derivation of Stirling's Approximation of $n!$, George Marsaglia and John C. W. Marsaglia, 826

A New Series Representation for $\zeta(3)$, John A. Ewell, 219

New Topologies from Old via Ideals, Dragan Janković and T. R. Hamlett, 295

A Note on Defining Conditional Probability, Arthur Ramer, 336

An Old Max-Min Problem Revisited, Mary Embry-Wardrop, 421

A One-Sentence Proof That Every Prime $p \equiv 1 (\mathrm{mod} 4)$ Is a Sum of Two Squares, D. Zagier, 144

Open Problems in Grid Labeling, Joseph A. Gallian, 133

Overdetermined Systems of Linear Equations, Gareth Williams, 511

Polynomials with Large Numbers of Prime Values, Betty Garrison, 316

Has Progress in Mathematics Slowed Down?, Paul R. Halmos, 561

Random Intervals, Joyce Justicz, Edward R. Scheinerman, and Peter M. Winkler, 881

Random Triangles, Joe Whittaker, 228

A Rearrangement Inequality and the Permutahedron, A. Vince, 319

Relations Between $\omega(n)$ and $\Omega(n)$, Armel Mercier, 503

Remainder Estimates in Taylor's Theorem, G. B. Folland, 233

The Remainder in Taylor's Formula, Esteban I. Poffald, 205

A Remark on the Shortest Confidence Interval of a Normal Mean, Robert Bartoszyński and Wai Chan, 415

Representing Binomial Coefficients as Sums of Squares, Andrew Granville and Yiliang Zhu, 486

Rouché's Theorem for Polynomials, Michael Filaseta, 834

Separable Extensions and Diagonalizability, Fred Richman, 395

The σ-Game and Cellular Automata, Klaus Sutner, 24

A Simple Algorithm for Finding Eigenvalues and Eigenvectors for 2×2 Matrices, Tyre A. Newton, 57

A Simple Derivation of the Maclaurin Series for Sine and Cosine, Deng Bo, 836

Singular Values and the Spectral Theorem, K. Hoechsmann, 413

The Solution of Certain Integral Equations by Means of Operators of Arbitrary Order, Bertram Ross and Baldev K. Sachdeva, 498

Some Models of Plane Geometries, Branko Grünbaum and Jan Mycielski, 839

The Square Pyramid Puzzle, W. S. Anglin, 120

Stability Properties of Geometric Inequalities, H. Groemer, 382

On Stability of Invariant Subspaces of Matrices, A. C. M. Ran and L. Rodman, 809

The Sum of Absolute Deviations and the Sum of Squared Deviations, David L. Farnsworth, 911

Symmetry in Logos and Hubcaps, Joseph A. Gallian, 235

A Theorem on the Numerators of the Bernoulli Numbers, Kurt Girstmair, 136

Toetjes, Tomás Feder, 785

Two Theorems on Singular Values and Eigenvalues, Roy Mathias, 47

Two Theorems, Two Sculptures, Two Posters, Helaman Ferguson, 589

Understanding Vector Fields, C. R. Curjel, 524

A Visit to the Newtonian N-body Problem via Elementary Complex Variables, Donald G. Saari, 105

When Quasinormal Implies Normal, Dean Hickerson, Sherman Stein, and Kenya Yamaoka, 514

A Whirlwind Tour of Computational Geometry, Ron Graham and Frances Yao, 687

The Width of a Rose Petal, S. C. Althoen and M. F. Wyneken, 907

AUTHOR INDEX

Aczél, J., Functional Equations and L'Hôpital's Rule in an Exact Poisson Derivation, 423

Althoen, S. C. and M. F. Wyneken, The Width of a Rose Petal, 907

Anglin, W. S., The Square Pyramid Puzzle, 120

Bartoszyński, Robert and Wai Chan, A Remark on the Shortest Confidence Interval of a Normal Mean, 415

Beck, Anatole, Greed Is (Sometimes) Not Enough, 289

Bender, Edward A., Fred Kochman, and Douglas B. West, Adding up to Powers, 139

Berger, Marcel, Convexity, 650

Beukers, Frits, Jean-Paul Bézivin, and Phillipe Robba, An Alternative Proof of the Lindemann-Weierstrass Theorem, 193

Bézivin, Jean-Paul, Frits Beukers, and Phillipe Robba, An Alternative Proof of the Lindemann-Weierstrass Theorem, 193

Bloom, Stephen L., Zoltán Ésik, and Ernest G. Manes, A Cayley Theorem for Boolean Algebras, 831

Bo, Deng, A Simple Derivation of the Maclaurin Series for Sine and Cosine, 836

Bollinger, Richard and Charles Burchard, Lucas's Theorem and Some Related Results for Extended Pascal Triangles, 198

Boyd, David W., The Diophantine Equation $x^2 + y^m = z^{2n}$, Addendum, 411

Burchard, Charles and Richard Bollinger, Lucas's Theorem and Some Related Results for Extended Pascal Triangles, 198

Burgess, C. E., Continuous Functions and Connected Graphs, 337

Byrnes, J. S. and Donald J. Newman, The L^4 Norm of a Polynomial with Coefficients ± 1, 42

Cannonito, F. B., On Inequivalent Group Extensions, 317

Chan, Wai and Robert Bartoszynski, A Remark on the Shortest Confidence Interval of a Normal Mean, 415

Chern, Shiing-Shen, What Is Geometry?, 679

Cohen, Regina B., Set Isometries and Their Extensions in the Max Metric, 795

Curjel, C. R., Understanding Vector Fields, 524

De Groen, Pieter P. N., A Counterexample on Vector Norms and the Subordinate Matrix Norms, 406

Desbrow, D., On the Irrationality of π^2, 903

Dubeau, François, Cauchy-Bunyakowski-Schwarz Inequality Revisited, 419

Eisele, P. and K. P. Hadeler, Game of Cards, Dynamical Systems, and a Characterization of the Floor and Ceiling Functions, 466

Embry-Wardrop, Mary, An Old Max-Min Problem Revisited, 421

Ésik, Zoltán, Stephen L. Bloom, and Ernest G. Manes, A Cayley Theorem for Boolean Algebras, 831

Ewell, John A., A New Series Representation for $\zeta(3)$, 219

Farnsworth, David L., The Sum of Absolute Deviations and the Sum of Squared Deviations, 911

Feder, Tomás, Toetjes, 785

Ferguson, Helaman, Two Theorems, Two Sculptures, Two Posters, 589

Filaseta, Michael, Rouché's Theorem for Polynomials, 834

Folland, G. B., Remainder Estimates in Taylor's Theorem, 233

Friedberg, Stephen H., Extending the Principal Axis Theorem to Fields Other Than $\mathbf{R}$, 147

Gallian, Joseph A., Open Problems in Grid Labeling, 133

Gallian, Joseph A., Symmetry in Logos and Hubcaps, 235

Garrison, Betty, Polynomials with Large Numbers of Prime Values, 316

Giblin, Peter J. and Donal B O'Shea, The Bitangent Sphere Problem, 5

Girstmair, Kurt, A Theorem on the Numerators of the Bernoulli Numbers, 136

Gorni, Gianluca, A Geometric Approach to l'Hôpital's Rule, 518

Graham, Ron and Frances Yao, A Whirlwind Tour of Computational Geometry, 687

Granville, Andrew and Yiliang Zhu, Representing Binomial Coefficients as Sums of Squares, 486

Groemer, H., Stability Properties of Geometric Inequalities, 382

Grünbaum, Branko and Jan Mycielski, Some Models of Plane Geometries, 839

Guy, Richard K., A Guessing Game of Bill Sands, and Bernardo Recamán's Barranca, 314

Hadeler, K. P. and P. Eisele, Game of Cards, Dynamical Systems, and a Characterization of the Floor and Ceiling Functions, 466

Halmos, Paul R., Has Progress in Mathematics Slowed Down?, 561

Hamlett, T. R. and Dragan Janković, New Topologies from Old via Ideals, 295

Hickerson, Dean, Sherman Stein and Kenya Yamaoka, When Quasinormal Implies Normal, 514

Hirschhorn, M. D., A Birthday Present for Ramanujan, 398

Hoechsmann, K., Singular Values and the Spectral Theorem, 413

Hoffman, David and William H. Meeks, III, Minimal Surfaces Based on the Catenoid, 702

Hungerford, Thomas W., A Counterexample in Galois Theory, 54

Jamieson, M. J., Fibonnaci Numbers and Aitken Sequences Revisited, 829

Janković, Dragan, and T. R. Hamlett, New Topologies from Old via Ideals, 295

Justicz, Joyce, Edward R. Scheinerman, and Peter M. Winkler, Random Intervals, 881

Kochman, Fred, Edward A. Bender, and Douglas B. West, Adding up to Powers, 139

Kopel, Daniel and Michael Schramm, A New Extension of the Derivative, 230

Krom, Melven and Myren Krom, More on Mortality, 37

Krom, Myren and Melven Krom, More on Mortality, 37

Kupitz, Yaakov S. and Micha A. Perles, A Condition for Flatness of Curves in R^n, 401

Lax, Peter D., The Ergodic Character of Sequences of Pedal Triangles, 377

Leslie, Robert A., The l_2 Matrix Norm and the Spectral Radius, 339

London, R. R. and H. P. Rogosinski, Decomposition Theory in the Teaching of Elementary Linear Algebra, 478

Manes, Ernest G., Stephen L. Bloom, and Zoltán Ésik, A Cayley Theorem for Boolean Algebras, 831

Marsaglia, George and John C. W. Marsaglia, A New Derivation of Stirling's Approximation of $n!$, 826

Marsaglia, John C. W. and George Marsaglia, A New Derivation of Stirling's Approximation of $n!$, 826

Martin, Gary A., For Your Eyes Only, 418

Mathews, Jerold, Gear Trains and Continued Fractions, 505

Mathias, Roy, Two Theorems on Singular Values and Eigenvalues, 47

Meeks, III, William H. and David Hoffman, Minimal Surfaces Based on the Catenoid, 702

Mercier, Armel, Relations Between $\omega(n)$ and $\Omega(n)$, 503

Meyer, Kenneth R., The Geometry of Harmonic Oscillators, 457

Minda, David, The Dirichlet Problem for a Disk, 220

Mirollo, R. E. and S. H. Strogatz, Integral Representation of a Finite Spike, 901

Morton, Patrick, Musings on the Prime Divisors of Arithmetic Sequences, 323

Mycielski, Jan and Branko Grünbaum, Some Models of Plane Geometries, 839

Newman, Donald J. and J. S. Byrnes, The L^4 Norm of a Polynomial with Coefficients ± 1, 42

Newton, Tyre A., A Simple Algorithm for Finding Eigenvalues and Eigenvectors for 2×2 Matrices, 57

Nitsche, Johannes C. C., Isoptic Characterization of a Circle (Proof of a Conjecture of M. S. Klamkin), 45

O'Shea, Donal B. and Peter J. Giblin, The Bitangent Sphere Problem, 5

Osserman, Robert, Curvature in the Eighties, 731

Perles, Micha A. and Yaakov S. Kupitz, A Condition for Flatness of Curves in R^n, 401

Petkovšek, Marko, Ambiguous Numbers are Dense, 408

Pillers, Carolyn and Tim Robertson, A Geometric Regularity From Order-Restricted Statistical Inference, 496

Poffald, Esteban I., The Remainder in Taylor's Formula, 205

Ramer, Arthur, A Note on Defining Conditional Probability, 336

Ran, A. C. M. and L. Rodman, On Stability of Invariant Subspaces of Matrices, 809

Richman, Fred, Separable Extensions and Diagonalizability, 395

Richter, R. Bruce and William P. Wardlaw, Diagonalization over Commutative Rings, 223

Robba, Philippe, F. Beukers, and J.-P. Bézivin, An Alternative Proof of the Lindemann-Weiestrass Theorem, 193

Robertson, Tim and Carolyn Pillers, A Geometric Regularity From Order-Restricted Statistical Inference, 496

Rodman, L. and A. C. M. Ran, On Stability of Invariant Subspaces of Matrices, 809

Rogosinski, H. P. and R. R. London, Decomposition Theory in the Teaching of Elementary Linear Algebra, 478

Ross, Bertram and Baldev K. Sachdeva, The Solution of Certain Integral Equations by Means of Operators of Arbitrary Order, 498

Saari, Donald G., A Visit to the Newtonian N-body Problem via Elementary Complex Variables, 105

Sachdeva, Baldev K. and Bertram Ross, The Solution of Certain Integral Equations by Means of Operators of Arbitrary Order, 498

Santa Coloma, Teresa L. and Ralph P. Tucci, A Decomposition Theorem for Finite Automata, 51

Scheinerman, Edward R., Joyce Justicz and Peter M. Winkler, Random Intervals, 881

Schramm, Michael and Daniel Kopel, A New Extension of the Derivative, 230

Schwarz, Gideon, The Dark Side of the Möbius Strip, 890

Senechal, Marjorie, Finding the Finite Groups of Symmetries of the Sphere, 329

Snapper, Ernst, Inverse Functions and their Derivatives, 144

Stein, Sherman, Dean Hickerson, and Kenya Yamaoka, When Quasinormal Implies Normal, 514

Strogatz, S. H. and R. E. Mirollo, Integral Representation of a Finite Spike, 901

Sutner, Klaus, The σ-Game and Cellular Automata, 24

Swan, Richard G., Another Proof of the Quadratic Reciprocity Theorem?, 138

Thomassen, Carsten, A Link Between the Jordan Curve Theorem and the Kuratowski Planarity Criterion, 216

Thurston, William P., Conway's Tiling Groups, 757

Tucci, Ralph P. and Teresa L. Santa Coloma, A Decomposition Theorem for Finite Automata, 51

Ungar, Peter, Mixing Property of the Pedal Mapping, 898

Van Daele, A., The Lebesgue Integral Without Measure Theory, 912

Van Den Essen, Arno, Magic Squares and Linear Algebra, 60

Vince, A., A Rearrangement Inequality and the Permutahedron, 319

Wagon, Stan, The Euclidean Algorithm Strikes Again, 125

Wardlaw, William P. and R. Bruce Richter, Diagonalization over Commutative Rings, 223

Weinstock, Robert, Elementary Evaluations of $\int_0^\infty e^{-x^2}dx$, $\int_0^\infty \cos x^2 dx$, and $\int_0^\infty \sin x^2 dx$, 39

West, Douglas B., Edward A. Bender, and Fred Kochman, Adding up to Powers, 139

Whittaker, Joe, Random Triangles, 228

Williams, Gareth, Overdetermined Systems of Linear Equations, 511

Wingler, Eric, An Infinite Product Expansion for the Square Root Function, 836

Winkler, Peter M., Joyce Justicz, and Edward R. Scheinerman, Random Intervals, 881

Wyneken, M. F. and S. C. Althoen, The Width of a Rose Petal, 907

Yamaoka, Kenya, Dean Hickerson, and Sherman Stein, When Quasinormal Implies Normal, 514

Yao, Frances and Ron Graham, A Whirlwind Tour of Computational Geometry, 687

Zagier, D., A One-Sentence Proof That Every Prime $p \equiv 1 (\mathrm{mod}\ 4)$ Is a Sum of Two Squares, 144

Zhu, Yiliang and Andrew Granville, Representing Binomial Coefficients as Sums of Squares, 486

SUBJECT INDEX

This index uses the 1985 revision of the AMS's 1980 Mathematics Subject Classification scheme.

00-XX GENERAL

-01 Elementary exposition; textbooks C.R. CURJEL 524

A05 General Mathematics M. EMBRY-WARDROP 421

03-XX MATHEMATICAL LOGIC AND FOUNDATIONS

D35 Recursion theory; Undecidability and degrees of sets of sentences M. KROM and M. KROM 37

05-XX COMBINATORICS

A10 Classical combinatorial problems; factorials, binomial coefficients, combinatorial functions R. BOLLINGER and C. BURCHARD 198

A19 Classical combinatorial problems; combinatorial identities E. A. BENDER, F. KOCHMAN, and D. B. WEST 139

A99 Classical combinatorial problems; none of the above, but in this section R. K. GUY

06-XX ORDER, LATTICES, ORDERED ALGEBRAIC STRUCTURES

E20 Boolean algebras; Boolean rings S. L. BLOOM, Z. ÉSIK, and E. G. MANES 831

11-XX NUMBER THEORY

A15 Elementary number theory; power residues, reciprocity R. G. SWAN 138

A63 Elementary number theory; radix representation; digital problems M. PETKOVŠEK 408

B65 Sequences and sets; binomial coefficients; factorials; q-identities A. GRANVILLE and Y. ZHU 486

D41 Diophantine equations; higher degree equations; Fermat's equation D. W. BOYD 411

N32 Multiplicative number theory; primes represented by polynomials; other multiplicative structures of polynomial values B. GARRISON 316

P76 Additive number theory; partitions; partitions: congruences for partition functions M. D. HIRSCHHORN 398

12-XX FIELD THEORY AND POLYNOMIALS

F10 Field extensions; separable extensions, Galois theory T. W. HUNGERFORD 54 F. RICHMAN 395

13-XX COMMUTATIVE RINGS AND ALGEBRAS

A99 General commutative ring theory; None of the above, but in this section R. B. RICHTER and W. P. WARDLAW 223

15-XX LINEAR AND MULTILINEAR ALGEBRA; MATRIX THEORY

A03 Vector spaces, linear dependence, rank A. VAN DEN ESSEN 60

A18 Eigenvalues, singular values, and eigenvectors T. A. NEWTON 57

A33 Matrices over special rings (quaternions, finite fields, etc.) S. H. FRIEDBERG 147

A63 Quadratic and bilinear forms, inner products F. DUBEAU 419

20-XX GROUP THEORY AND GENERALIZATIONS

-01 Elementary exposition; textbooks D. HICKERSON, S. STEIN, and K. YAMAOKA 514

E22 Structure and classification of infinite or finite groups; extensions, wreath products, and other compositions F. B. CANNONITO 317

M35 Semigroups; semigroups in automata theory, linguistics, etc. T. L. SANTA COLOMA and R. P. TUCCI 51

26-XX REAL FUNCTIONS

-01 Elementary exposition; textbooks S. C. ALTHOEN and M. F. WYNEKEN 907

A06 Functions of one variable; one-variable calculus G. B. FOLLAND 233 G. GORNI 518

A24 Functions of one variable; differentiation: general theory, generalized derivatives, mean-value theorems E. I. POFFALD 205

28-XX MEASURE AND INTEGRATION

-01 Elementary exposition; textbooks A. VAN DAELE 912

34-XX ORDINARY DIFFERENTIAL EQUATIONS

-01 Elementary exposition; textbooks K. R. MEYER 457

39-XX FINITE DIFFERENCES AND FUNCTIONAL EQUATIONS

B20 Functional equations; equations for one unknown function of one variable, rank greater than one J. ACZÉL 423

40-XX SEQUENCES, SERIES, SUMMABILITY

B05 Multiple sequences and series DENG BO 836

51-XX GEOMETRY

-01 Elementary exposition; textbooks PETER D. LAX 377 BRANKO GRÜNBAUM and JAN MYCIELSKI 839

M20 Real and complex geometry; polyhedra and polytopes; regular figures, division of spaces C. PILLERS and T. ROBERTSON 496

52-XX CONVEX SETS AND RELATED GEOMETRIC TOPICS

A40 Inequalities and extremum problems H. GROEMER 382

54-XX GENERAL TOPOLOGY

C05 Maps and general types of spaces defined by maps; continuous maps C. E. BURGESS 337

60-XX PROBABILITY THEORY AND STOCHASTIC PROCESSES

D05 Geometric probability, stochastic geometry, random sets J. WHITTAKER 228

E15 Distribution theory; distributions: general theory R. BARTOSZYŃSKI and W. CHAN 415

62-XX STATISTICS

A99 Foundations; None of the above, but in this section D. L. FARNSWORTH 911

68-XX COMPUTER SCIENCE

Q80 Theory of computing; tesselation automata, iterative arrays, cellular structure KLAUS SUTNER 24

REVIEWS BY TITLE

Names of authors are in ordinary type; those of reviewers in capitals.

Basic Hypergeometric Series and Applications, Nathan J. Fine, RANJAN ROY, 82

Complex Analysis, Joseph Bak and Donald J. Newman, LAWRENCE ZALCMAN, 262

A Course in Constructive Algebra, R. Mines, F. Richman, and W. Ruitenberg, MICHAEL J. BEESON, 357

The Emperor's New Mind, Roger Penrose, MICHAEL BARR, 938

A Formalization of Set Theory without Variables, Alfred Tarski and Steven Givant, LISL GAAL, 258

Fractals Everywhere, Michael Barnsley, ANTHONY BARCELLOS, 266

Generatingfunctionology, Herbert S. Wilf, RICHARD P. STANLEY, 864

Innumeracy: Mathematical Illiteracy and Its Consequences, John Allen Paulos, LISA J. EVERED, 88

Invitation to Complex Analysis, Ralph Philip Boas, LAWRENCE ZALCMAN, 262

Lattice Points, Paul Erdős, Peter M. Gruber, and Joseph Hammer, WILLIAM FIREY, 633

Learning Discrete Mathematics With ISETL, Nancy Baxter, Ed Dubinsky, and Gary Levin, DONALD L. MEUNCH, 861

The Mathematics of Plato's Academy. A New Reconstruction, D. H. Fowler, RICHARD K. GUY, 440

Perspectives of Nonlinear Dynamics, E. Atlee Jackson, JOHN M. GUCKENHEIMER, 544

*Representations of *-Algebras, Locally Compact Groups, and Banach *-Algebraic Bundles,* J. M. G. Fell and R. S. Doran, MARC A. RIEFFEL, 167

The Turing Omnibus: 61 Excursions in Computer Science, A. K. Dewdney, MARK BRIDGER, 355

REVIEWS BY AUTHOR

Names of authors are in ordinary type; those of reviewers in capitals.

Bak, Joseph and Donald J. Newman, *Complex Analysis,* LAWRENCE ZALCMAN, 262

Barnsley, Michael, *Fractals Everywhere,* ANTHONY BARCELLOS, 266

Baxter, Nancy, Ed Dubinsky, and Gary Levin, *Learning Discrete Mathematics With ISETL,* DONALD L. MEUNCH, 861

Boas, Ralph Philip, *Invitation to Complex Analysis,* LAWRENCE ZALCMAN, 262

Dewdney, A. K., *The Turing Omnibus: 61 Excursions in Computer Science,* MARK BRIDGER, 355

Doran, R. S. and J. M. G. Fell, *Representations of *-Algebras, Locally Compact Groups, and Banach *-Algebraic Bundles,* MARC A. RIEFFEL, 167

Dubinsky, Ed, Nancy Baxter, and Gary Levin, *Learning Discrete Mathematics With ISETL,* DONALD L. MEUNCH, 861

Erdős, Paul, Peter M. Gruber and Joseph Hammer, *Lattice Points,* WILLIAM FIREY, 633

Fell, J. M. G. and R. S. Doran, *Representations of *-Algebras, Locally Compact Groups, and Banach *-Algebraic Bundles,* MARC A. RIEFFEL, 167

Fine, Nathan J., *Basic Hypergeometric Series and Applications,* RANJAN ROY, 82

Fowler, D. H., *The Mathematics of Plato's Academy. A New Reconstruction,* RICHARD K. GUY, 440

Givant, Steven and Alfred Tarski, *A Formalization of Set Theory without Variables,* LISL GAAL, 258

Gruber, Peter M., Paul Erdős and Joseph Hammer, *Lattice Points,* WILLIAM FIREY, 633

Hammer, Joseph, Paul Erdős and Peter M. Gruber, *Lattice Points,* WILLIAM FIREY, 633

Jackson, E. Atlee, *Perspectives of Nonlinear Dynamics,* JOHN M. GUCKENHEIMER, 544

Levin, Gary, Nancy Baxter, and Ed Dubinsky, *Learning Discrete Mathematics With ISETL,* DONALD L. MEUNCH, 861

Mines, R., F. Richman, and W. Ruitenberg, *A Course in Constructive Algebra,* MICHAEL J. BEESON, 357

Newman, Donald J. and Joseph Bak, *Complex Analysis,* LAWRENCE ZALCMAN, 262

Paulos, John Allen, *Innumeracy: Mathematical Illiteracy and Its Consequences,* LISA J. EVERED, 88

Penrose, Roger, *The Emperor's New Mind,* MICHAEL BARR, 938

Richman, F., R. Mines, and W. Ruitenberg, *A Course in Constructive Algebra,* MICHAEL J. BEESON, 357

Ruitenberg, W., R. Mines, and F. Richman, *A Course in Constructive Algebra,* MICHAEL J. BEESON, 357

Tarski, Alfred and Steven Givant, *A Formalization of Set Theory without Variables,* LISL GAAL, 258

Wilf, Herbert S., *Generatingfunctionology,* RICHARD P. STANLEY, 864

INDEX OF SPECIAL TOPICS

Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to Leon Henkin 1

Special Issue on Geometry (October 1990)

Convexity, Marcel Berger 650

What is Geometry?, Shiing-Shen Chern 679

A Whirlwind Tour of Computational Geometry, Ron Graham and Frances Yao 687

Minimal Surfaces Based on the Catenoid, David Hoffman and William H. Meeks, III 702

Curvature in the Eighties, Robert Osserman 731

Conway's Tiling Groups, William P.Thurston 757

SOLUTIONS

Numbers in boldface refer to problems; those in lightface to pages.

E 3094 64
E 3116 65
E 3167 918
E 3207 67
E 3222 152
E 3224 153
E 3225 154
E 3229 70
E 3236 529
E 3239 531
E 3251 612
E 3253 921
E 3254 922
E 3255 848
E 3256 73
E 3259 156
E 3260 74
E 3262 240
E 3264 157
E 3265 242
E 3273 159
E 3275 160
E 3276 244
E 3277 245
E 3278 247
E 3279 75
E 3280 160
E 3281 924
E 3282 925
E 3283 849
E 3284 248
E 3285 532
E 3286 249
E 3287 343
E 3288 344
E 3289 346
E 3290 428
E 3291 346
E 3294 347
E 3295 161
E 3297 250
E 3298 430
E 3300 251
E 3301 431
E 3302 432
E 3303 348
E 3304 534
E 3305 349
E 3309 614
E 3310 615
E 3311 617
E 3313 850
E 3314 851
E 3316 852
E 3317 618
E 3318 853
E 3319 854
E 3320 619
E 3321 620
E 3322 621
E 3325 855
E 3331 926
E 3334 856
E 3335 927
E 3336 856
E 3337 927
6478 858
6526 162
6539 253
6546 433
6552 622
6555 351
6557 535
6558 434
6559 76
6564 78
6565 80
6566 353
6567 165
6569 166
6570 254
6572 537
6573 540
6575 256
6576 930
6577 436
6578 542
6579 625
6580 437
6581 626
6582 438
6584 630
6585 859
6586 932
6589 933
6590 934
6600 935

PROBLEMS PROPOSED

Allegretto, W. and A. Meir 76
Alper, Theodore 857
Bòna, Miklòs 240
Bassali, W. A. 151
Bege, Antal 528
Berndt, Bruce C. 528
Blodgett, Robert J. 240
Boese, G. 252
Bowman, Douglas 857
Briggs, W. E. 350
Brown, Gavin 151
Chapman, R. J. 239
Chen, Ji and Cheng-hui Lo 611
Clark, Dean S. 64
Cook, Matthew, W. Janous, and M. E. Kuczma 428
Cossi, Ernesto Bruno 162
Dalang, Robert C. 161
Davis, James A. and J. Van Bowen 612
Dearden, Bruce G. and Michael B. Gregory 252
Diamond, H. G. 343
Doster, David 528
Erdős, Paul 847
Evans, Ronald 432
Ferrer, Jesús and Juan V. Savall 848
Fine, N. J. 63
Fine, N. J. 239
Floyd, Robert W. 611
Freyd, Peter 916
Fukuta, Jiro 150
Galvin, Fred, G. Muraz, and P. Szeptycki 433
Garsia, A. M. and B. A. Sethuraman 151
Gauchman, Hillel 529
Golomb, Solomon W. 427
Gregory, Michael B. and Bruce G. Dearden 252
Gurarie, David 342
Gurevič, Reuven 847
Haible, Bruno 529
Holshouser, Arthur L. 75
Israel, Robert B. 428
Janous, Walther, M. Cook, and M. E. Kuczma 428
Kieffer, John 150
Kimberling, Clark 848
Knuth, Donald E. and Boris Pittel 916
Kreczner, Robert 930
Kuczma, Marcin E., M. Cook, and W. Janous 428
Kupka, Joseph 612
Lenstra, H. W 64
Leonard, I. E. 150
Letac, Gérard 929
Lo, Cheng-hui and Ji Chen 611
MacKinnon, Nick 342
McNaughton, Robert 917
Meir, A. and W. Allegretto 76
Melter, Robert A. 350
Monsky, Paul 622
Muraz, Gilbert, P. Szeptycki, and F. Galvin 433
Myerson, Gerald 917
Pagus, David 350
Philipp, Stan 622
Pittel, Boris and Donald E. Knuth 916
Poonen, Bjorn 63
Porta, Horacio 433
Pouryoussefi, Farhoud 240
Robinson, Geoffrey R. 343
Rogerson, Peter 929
Ross, K. A. 162
Savall, Juan V. and Jesús Ferrer 848
Schmidt, F. 535
Schuster, Eugene F. 427
Sethuraman, B. A. and A. M. Garsia 151
Shallit, Jeffrey 848
Silverman, Joseph H. 428
Song-Qing, Mo 529
Steele, J. Michael 535
Stein, Alan H. 611
Strauss, Nicholas 252
Sturm, Jacob 427
Szeptycki, Pawel, G. Muraz, and F. Galvin 433
Thurston, Hugh 342
Tomescu, Ioan 916
Totaro, Burt J. 612
Vaaler, Jeffrey 239
Van Bowen, J. and James Davis 612
Vanden Eynden, Charles 847
Wardlaw, William P. 63
Waterhouse, William C. 76
Wepsic, Eric 917
West, Douglas B. 428
Wilf, Herbert S. 621
Zeamer, R. W. 535

PROBLEMS SOLVED

Alsholm, Preben 70
Andersen, K. F. 853, 855
Arias-de-Reyna, Juan, M. M. Day, and N. T. Peck 433
Ashbaugh, Mark 619
Bailey, Duane W. 251, 621
Binz, J. C. 922
Bivens, Irl C. 428
Boersma, J. 539
Bondesen, Aage and Douglas B. West 923
Boros, Endre and András Prékopa 152
Borwein, J. M. and P. B. Borwein 927
Breen, Mike, R. Douglass, and R. B. Eggleton 432
Brown, Kevin 933
Callan, David 250, 349, 848, 614
Cook, Matthew 918
Day, Mahlon M., J. Arias-de-Reyna, and N. T. Peck 433
Delany, Jim and Kempton Huehn 153
Douglass, Roger, M. Breen, and R. B. Eggleton 432
Edgar, G. A. and Klaus Zacharias 157
Edgar, Gerald A. and Lee A. Rubel 165
Eggleton, Roger B., R. Douglass, and M. Breen 432
Estes, Dennis R. 249
Ferrer, Jesús 347, 928, 617
Ferrer, Jesús and S. M. Gagola, Jr. 73
Florida State University Mail Room Problem Group 436
Franco, Zachary 344
Frohliger, John A. 543
Fukuta, Jiro 530
Gagola, Jr., Stephen M. and Jesús Ferrer 73
Galvin, Fred 166
George Washington University Problem Group 152
Gray, Alfred 625
Grossman, Nathaniel 935
Grosswald, Emil 241
Guggenheimer, H. 431
Harris Kwong, Y. H. 850
Henley, Douglas and Joseph E. McCanna 81
Hensley, Douglas A. and Claudia A. Spiro 253
Hipp, Christian 930
Huehn, Kempton and Jim Delany 153
Ilani, Ishai 65
Israel, Robert B. 158
Jager, Thomas 245
Jagers, A. A. 246, 438, 934
Janous, Walther 532, 921, 922, 925
Kappus, Hans 927
Kastanas, Ilias G. 854
Ki, Ha-Seo 79
Kim, Byeong Soo 346
Knuth, Donald E. 256, 348, 629
Kuczma, Marcin E. 618, 851
Lau, Kee-Wai 161, 244, 345
Levin, Stewart A. 857
Lossers, O. P. 76, 162, 343, 436, 536, 623, 931
Macdonald, I. G. 613
Martins, Luiz Felipe 856
Mattics, L. E. 241, 351, 623
McCanna, Joseph and Douglas Henley 81
Montgomery, Peter 250
Mortini, Raymond 855
Nieto, José Heber 160, 348
Nijenhuis, Albert 159, 246, 853
Osterlé, Joseph 537
Oursler, C. C. 534
Passell, Nicholas 353
Peck, N. Tenney, J. Arias-de-Reyna, and M. M. Day 433
Pederson, Allan 849
Philipp, Stan 155
Portnoy, Stephen 540
Powell, Barry 434
Powers, Glenn 251
Prékopa, András and Endre Boros 152
Pryor, Doug 243
Rabinowitz, Stanley 75
Raychev, Emil 156
Richberg, Rolf 626
Riese, Adam 855
Robinson, David G. 69
Rogers, C. A. 64, 858
Rubel, L. A. 255
Rubel, Lee A. and Gerald A. Edgar 165
Ruzsa, Imre Z. 615
Selfridge, J. L. 437
Shippensburg University Mathematical Problem Solving Group 247
Spiro, Claudia A. and Douglas A. Henley 253
Stafford, Richard M. 630
Starbird, Thomas W. 932
Steelman, John Henry 74, 160, 247, 429, 430, 436, 620, 924
Stong, Richard 248, 429
Students of the 1987 Mathematical Olympiad Program 68

University of South Alabama Problem Group 67
Virginia Commonwealth University Algebra Students 532
Vowe, Michael 856
West, Douglas B. and Aage Bondesen 923
Western Maryland College Problems Group 859
Wilker, J. B. 251
Williams, Robert C. 346
Wong, Yan-Loi 246
Woo, Jungmin 927
Zacharias, Klaus and G. A. Edgar 157
Zulauf, A. 615
Zwicker, William S. 926

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 97, Number 10 December 1990

Contents

(ISSN 0002-9890)

ARTICLES

Random Intervals
JOYCE JUSTICZ, EDWARD R. SCHEINERMAN, AND PETER M. WINKLER 881

The Dark Side of the Moebius Strip GIDEON SCHWARZ 890

NOTES

Mixing Property of the Pedal Mapping PETER UNGAR 898

Integral Representation of a Finite Spike R. E. MIROLLO AND S. H. STROGATZ 901

On the Irrationality of π^2 D. DESBROW 903

THE TEACHING OF MATHEMATICS

The Width of a Rose Petal S. C. ALTHOEN AND M. F. WYNEKEN 907

The Sum of Absolute Deviations and the Sum of Squared Deviations DAVID L. FARNSWORTH 911

The Lebesgue Integral Without Measure Theory A. VAN DAELE 912

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions 916

Advanced Problems and Solutions 929

REVIEWS

The Emperor's New Mind. By Roger Penrose MICHAEL BARR 938

TELEGRAPHIC REVIEWS 943

INDEX TO VOLUME 97 949

NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Macalester College, St. Paul, MN 55105. Other shroter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.

EDITOR

HERBERT S. WILF

ASSOCIATE EDITORS

PAUL T. BATEMAN
DONNA BEERS
DENNIS DETURCK
HAROLD G. DIAMOND
JOHN DIXON
J. H. EWING
RICHARD GUY
DAVID J. HALLENBECK
PAUL R. HALMOS
MELVIN HENRIKSEN
JOAN P. HUTCHINSON
WILLIAM M. KANTOR
JOSEPH KONHAUSER
RICHARD LIBERA
LEE A. RUBEL
RODICA SIMION
ANITA E. SOLOW
JOEL SPENCER
LYNN A. STEEN
KENNETH B. STOLARSKY
STAN WAGON
DOUGLAS B. WEST

Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $30 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $60. Student and unemployed members receive a 66% dues discount; emeritus members receive a 50% dues discount; and new members receive a 40% dues discount for the first two years of membership.) The nonmember/library subscription price is $120 per year.

Copyright © by the Mathematical Association of America (Incorporated), 1990, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source.

Second class postage paid at Washington, DC, and additional mailing offices.

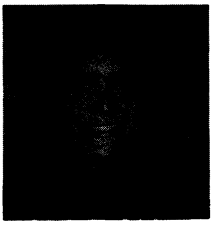
Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

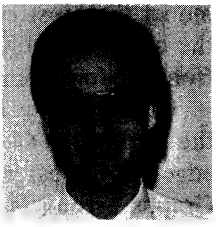
Random Intervals

JOYCE JUSTICZ, EDWARD R. SCHEINERMAN, AND PETER M. WINKLER

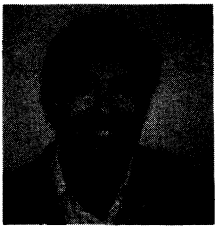
JOYCE JUSTICZ is a graduate student in mathematics at Emory University, where she received her bachelor's degree (as valedictorian) in 1985.



EDWARD SCHEINERMAN is an associate professor in the Department of Mathematical Sciences of the Johns Hopkins University. He received his Sc.B. from Brown University in 1980 and his Ph.D. from Princeton in 1984.



PETER WINKLER is Professor of Mathematics and Computer Science at Emory University, and manager of the Research Group in Mathematics and Theoretical Computer Science at Bellcore. His primary mathematical interests are combinatorics, logic and probabilistic methods.



Abstract. Fix a large number n and let $\{x_1, y_1, \ldots, x_n, y_n\}$ be $2n$ points chosen independently from some fixed continuous probability distribution on the real line. Each pair $\{x_i, y_i\}$ determines a *random interval* $[\min(x_i, y_i), \max(x_i, y_i)]$. We examine the structure of the resulting family of intervals, and in particular answer the following questions: how large a subcollection of pairwise disjoint intervals can one expect to find? And, what is the probability that there is an interval in the family which intersects all the others?

Prelude. Before beginning, we invite the reader to test his intuition on the following problem (it won't be any worse than ours was!). The numbers from 1 to $2n$, with n large, are paired at random, each pair being regarded as the endpoints of a real interval. What is the probability that among these n intervals there is one which meets all the others?

1. Introduction: Random Intervals. In studying any type of combinatorial structure it is useful, and sometimes quite illuminating, to have models for "random" structures of that type. The most famous example is the random graph model of

Erdös and Rényi [2] in which edges are chosen independently with fixed probability. Random graph theory now forms a substantial subject of study in itself.

The combinatorial structures with which we concern ourselves here are those based on intersection properties of families of intervals. The most obvious definition of a "random interval" is the interval between two random numbers x and y; for convenience we denote that interval by $[x, y]$ even when $y < x$, so that by definition $[x, y] = [y, x]$.

We create a family F of random intervals in the following way: fix a number n and a continuous probability distribution on the real line. Choose values $x_1, y_1, \ldots, x_n, y_n$ independently and let F consist of intervals $I_1, \ldots, I_n$ where $I_j = [x_j, y_j]$.

Since intersection properties of F depend only on the order of the points $x_1, \ldots, y_n$, it is clear that the model is insensitive to the choice of distribution. In fact, an equivalent discrete model is obtained by choosing at random one of the $(2n)!$ assignments of $x_1, \ldots, y_n$ to the integers from 1 to $2n$. In the continuous model the most convenient distribution seems to be the uniform distribution on the unit interval [0, 1]; in that case we may think of the intervals as having arisen from random points (x_i, y_i) in the unit square.

To any family F of intervals we may associate a graph $G = (V, E)$ as follows: the vertices of G correspond to the intervals of F, and two vertices constitute an edge of G just when their corresponding intervals have non-null intersection. Graphs arising in this fashion are called *interval graphs*; when F is a family of random intervals as described above, they are called *random interval graphs* [8].

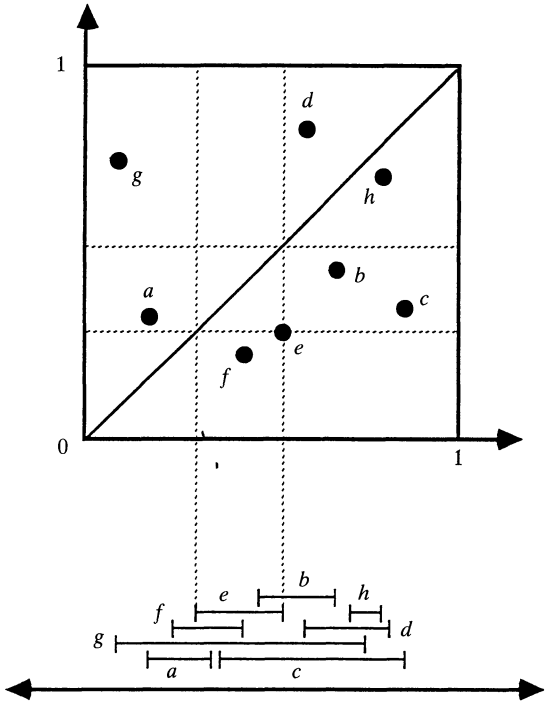


FIG. 1a. Eight random points and their intervals.

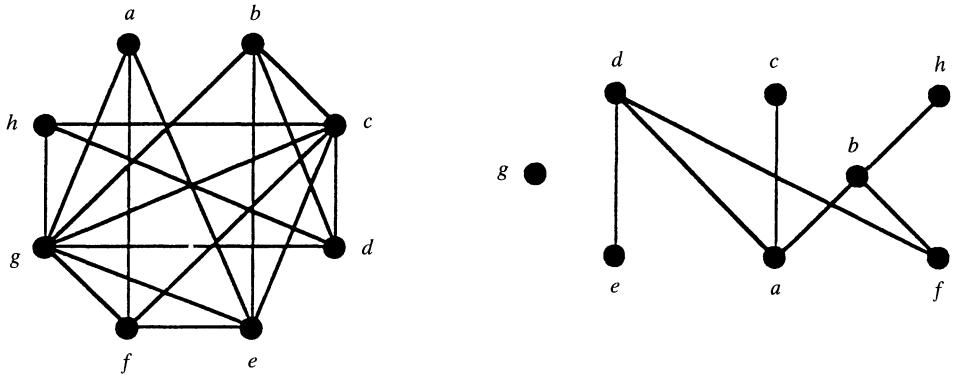


FIG. 1b. Random interval graph and interval order.

We preserve somewhat more of the properties of a family F of intervals by assigning to F a partially ordered set P; here the elements of P correspond to the intervals of F with $x \prec y$ just when the interval corresponding to x lies entirely to the left of the interval corresponding to y. Partially ordered sets arising in this fashion are called *interval orders*. Note that G may be obtained from P by declaring $\{x, y\}$ to be an edge of G whenever x and y are incomparable in P.

For further information about both interval graphs and interval orders, we refer the reader to Fishburn's excellent book [**4**].

FIGURE 1a depicts 8 random intervals (arising from 8 points chosen in the unit square) and FIGURE 1b depicts the corresponding interval graph and interval order.

2. Chains and Antichains.... or cliques and independent sets.

Given a collection of n random intervals, what is the size of a largest family of pairwise intersecting intervals? ... or pairwise disjoint intervals? In poset language we are asking for the sizes of the largest antichain (pairwise incomparable elements) and of the largest chain (pairwise comparable, and therefore totally ordered, elements). In graph language we seek the sizes of the largest clique (pairwise adjacent) and the largest independent set (pairwise nonadjacent).

The first question (pairwise intersecting) is easier to answer.

Intervals of reals satisfy the "Helly property," that is, every collection of pairwise-intersecting intervals has a nonempty intersection. Thus, when endpoints are uniformly distributed on $[0, 1]$, it is enough to find the point $x \in [0, 1]$ which is contained in the maximum number of intervals in the collection. The probability that an interval contains x is $1 - x^2 - (1 - x)^2 = 2x - 2x^2$. This is maximized when $x = 1/2$, and the expected number of intervals which contain $1/2$ is $n/2$. With a bit more work (the interested reader can consult [**8**]) one can obtain the following result:

THEOREM 1. *Let the random variable A_n denote the size of a largest set of pairwise intersecting intervals in a family of n random intervals. There exists a function $f(n)$ such that*

$$\lim_{n\to\infty} \frac{f(n)}{n} = 0$$

and

$$\lim_{n\to\infty} \Pr\left\{\frac{n}{2} - f(n) \leqslant A_n \leqslant \frac{n}{2} + f(n)\right\} = 1.$$

We turn now to the second question: What is the largest collection of pairwise non-intersecting intervals (longest chain) in a family of n random intervals? It turns out that this question is closely related to the following well-known problem, due to Stanislaw Ulam: What is the length of the longest increasing subsequence of a random permutation of $1, \ldots, n$?

Ulam's problem, posed in [**10**], was tackled by Hammersley in a famous paper [**6**] in which the answer was shown to be asymptotic to $c\sqrt{n}$ for some constant c. Later, a combination of the efforts of Schensted [**9**], Logan and Shepp [**7**] and Veršik and Kerov [**11**] determined that the constant is, in fact, 2.

Ulam's "random permutation" may be constructed by choosing n random points from the uniform distribution in the unit square, and numbering the points from left to right (i.e., according to their x-coordinates). The permutation arises in reading the label-numbers from bottom to top.

The points determine a partially ordered set Q under the product ordering, where $(x_i, y_i) \preceq_Q (x_j, y_j)$ just when $x_i \leqslant x_j$ and $y_i \leqslant y_j$. An increasing subsequence of the permutation corresponds exactly to a chain in Q; geometrically, a subset of the points extending from southwest to northeast in the square. Let us denote by X_n the random variable whose value is the size of the largest such subset.

As we have seen, the same set of points in the unit square determines a family of random intervals, but in the interval order P, we have the stronger condition $(x_i, y_i) \prec_P (x_j, y_j)$ iff $\max(x_i, y_i) < \min(x_j, y_j)$. Let us denote by Y_n the size of a largest chain in P; it is the asymptotic behavior of Y_n which we wish to determine. Since every chain of P is a chain of Q, we have $Y_n \leqslant X_n$.

For example, consider the point **e** in FIGURE 1a. The points greater than **e** in Q are those to the northeast (namely, **b**, **c**, **d** and **h**). We can also describe geometrically the points which lie above **e** in P. Draw a vertical and horizontal line through the point **e**. Also, draw a vertical and horizontal line through **e**'s reflection across the positively sloped diagonal of the square. These four lines divide the unit square into 9 rectangles. (See FIGURE 1a.) The upper right-hand square contains the points which correspond to intervals to the right of **e** (namely, **d** and **h**).

Hammersley [**6**] used the method of subadditivity to prove that $X_n/\sqrt{n}$ approaches a constant in probability, and in fact a nearly identical argument can be made to achieve the same result for $Y_n/\sqrt{n}$. (See, for example, Bolobás and Winkler [**1**] where Hammersley's results are extended to higher dimension.)

In the case of Ulam's problem the actual value of c was obtained only with great difficulty, and in fact the values of the constants in higher dimensions remain unknown. Luckily, a special feature of the interval case allows us both to prove the existence of the constant and determine its value with relative ease. The difference is that in our interval problem a maximum-length chain can be built in "greedy" fashion from the bottom up, whereas the equivalent process in Ulam's problem fails by a constant factor, in the limit, to attain maximal length.

THEOREM 2. *Let Y_n denote the maximum number of pairwise disjoint intervals in a family of n random intervals. Then*

$$\lim_{n\to\infty} \frac{Y_n}{\sqrt{n}} = \frac{2}{\sqrt{\pi}}$$

in probability.

This means for every $\varepsilon > 0$, there exists an n_0 so that for all $n > n_0$,

$$\Pr\left\{\left|\frac{Y_n}{\sqrt{n}} - \frac{2}{\sqrt{\pi}}\right| \leqslant \varepsilon\right\} > 1 - \varepsilon.$$

Proof. Let us establish a Poisson process on the plane of density 1. We select an infinite chain $C = \{(u_1, v_1), (u_2, v_2), \ldots\}$ of points of the process in the following manner: (u_1, v_1) is the point in the positive quadrant which minimizes $\max(u_1, v_1)$, and thereafter, (u_k, v_k) is the point above (u_{k-1}, v_{k-1}) which minimizes $\max(u_k, v_k)$. In terms of intervals, C represents the chain built from the bottom by always selecting the interval with least possible upper endpoint; it is easily seen by induction that in any finite collection of intervals such a chain has maximum possible length.

Now, for any positive real s, the region $\{(x, y) : 0 \leqslant x, y \leqslant s\} = [0, s]^2$ is with probability exactly $\exp\{-s^2\}$ unoccupied by a point of the Poisson process. It follows that if S is the random variable whose value is $\max(x_1, y_1)$, then the mass density of S is given by the function

$$f(s) = \frac{d}{ds}(1 - e^{-s^2}) = 2se^{-s^2}$$

whose expected value is

$$\int_0^\infty 2s^2 e^{-s^2}\, ds = \int_0^\infty t^{1/2} e^{-t}\, dt = \Gamma(3/2) = \frac{\sqrt{\pi}}{2}.$$

The differences

$$\max(u_1, v_1) - 0, \max(u_2, v_2) - \max(u_1, v_1), \max(u_3, v_3) - \max(u_2, v_2), \ldots$$

are independent and identically distributed with mean $\sqrt{\pi}/2$. It follows from the law of large numbers that for any $\varepsilon > 0$,

$$(1 - \varepsilon)\frac{\sqrt{\pi}}{2} < \frac{\max(x_m, y_m)}{m} < (1 + \varepsilon)\frac{\sqrt{\pi}}{2}$$

with probability at least $1 - \varepsilon$, for every sufficiently large m.

Now let $r(n)$ be the least r such that $[0, r]^2$ contains exactly n points of the Poisson process; these points then determine a family of n random intervals, as described above, and we may therefore identify Y_n with the largest m such that (u_m, v_m) lies in the square $[0, r(n)]^2$. Since the Poisson process has density 1, we will have

$$(1 - \varepsilon)\sqrt{n} < u(n) < (1 + \varepsilon)\sqrt{n}$$

with probability at least $1 - \varepsilon$, for sufficiently large n.

Let $m_1 = \lfloor (1 - \varepsilon)(2/\sqrt{\pi})\sqrt{n} \rfloor$ and $m_2 = \lfloor (1 + \varepsilon)(2/\sqrt{\pi})\sqrt{n} \rfloor$; then for large enough n, we have that (u_{m_1}, v_{m_1}) will lie inside the square $[0, r(n)]^2$ and (u_{m_2}, v_{m_2}) outside the square, with probability at least $1 - \varepsilon$. We conclude that

$$(1 - \varepsilon)\frac{2}{\sqrt{\pi}} < \frac{Y_n}{\sqrt{n}} < (1 + \varepsilon)\frac{2}{\sqrt{\pi}}$$

with probability at least $1 - \varepsilon$, proving the theorem.

Note: The Poisson process enables us to place the random variables Y_n all in the same sample space; thus, with the help of the strong law of large numbers, we may

obtain the slightly stronger result that

$$\lim_{n\to\infty} \frac{Y_n}{\sqrt{n}} = \frac{2}{\sqrt{\pi}}$$

with probability 1.

3. A Matter of Degree. In a graph, the degree of a vertex x, denoted $d(x)$, is the number of edges incident with x. The minimum and maximum degrees of a graph are denoted by δ and Δ respectively. What can we say about the minimum and maximum degree of a random interval graph?

The minimum degree question is answered in [**8**]. We simply repeat the result here:

THEOREM 3. *If δ_n denotes the minimum degree of a vertex in the interval graph generated by n random intervals, then for fixed $k > 0$,*

$$\lim_{n\to\infty} \Pr\left(\delta_n \leqslant k\sqrt{n}\right) = 1 - \exp\{-k^2/2\}.$$

It follows that the average minimum degree approaches $\sqrt{n\pi/2}$.

The maximum degree question is more interesting, partly on account of the role played by a point of degree $n - 1$. Recall that the *diameter* of a graph is the maximum over all pairs (x, y) of vertices of the least number of edges in a path from x to y. It is easy to show that with probability approaching 1, the diameter of our random interval graph is 2 or 3; and, for an interval graph G, $\operatorname{diam}(G) \leqslant 2$ iff there is a vertex adjacent to all others. In fact, the following are equivalent for any collection of n intervals:

1. the interval graph G has a vertex of degree $n - 1$;
2. G has diameter at most 2;
3. the interval order P has an isolated point;
4. there is an interval in the family which meets all the others; and
5. there is an interval in the family which meets both $[u, v]$ and $[p, q]$, where $[u,v]$ is the interval with leftmost right endpoint and $[p, q]$ the interval with rightmost left endpoint.

We are thus moved to ask: What is the limiting probability that in a family of random intervals, there is an interval which intersects all the others?

One normally expects such limits to be either 0 or 1, and in fact in many classes of structures (such as the random graphs of Erdös and Rényi) there is a "0-1 Law" —in that case proved by Fagin [**3**]—which guarantees that first-order statements, such as "there is a vertex which is adjacent to all others," must have trivial limiting probability. Exceptions to this sort of behavior are sometimes quite startling, as in the famous "probleme de rencontres" in which the probability that a random permutation has no fixed point approaches $1/e$.

Here we shall obtain an answer which not only ruins a possible 0-1 Law for random interval graphs, but adds insult to injury in a surprising way.

Let us consider first the Poisson model, except that this time for convenience, let the process be of density n on $[0, 1]^2$. Let S be the minimum, over all points (x, y) of the process, of $\max\{x, y\}$; thus $S = \max\{u, v\}$ where $[u, v]$ is the interval with leftmost right endpoint mentioned above. Similarly let R be the position of the rightmost left endpoint. Then the "big interval," i.e., the interval which

intersects all others, will exist just when there is a point of the process in the union of the rectangles $[0, S] \times [R, 1]$ and $[R, 1] \times [0, X]$. This will occur with limiting probability

$$\lim_{n\to\infty}\left[\int_0^\infty\int_0^\infty e^{-2nxy}\frac{\partial^2}{\partial x\,\partial y}\left(1 - e^{-nx^2} - e^{-ny^2} + e^{-n(x^2+y^2)}\right)dx\,dy\right] = \frac{2}{3}.$$

It was when the authors asked themselves how fast the probability converged to $2/3$ that the startling truth emerged:

THEOREM 4. *For all $n > 1$, the probability that in a collection of n random intervals there is one which intersects all the others is exactly $2/3$.*

Proof. One can show, using the uniform model on $[0, 1]^2$, that the above probability is

$$1 - 4n(n-1)\int_0^1\int_0^{1-y} xy\left(1 - x^2 - y^2 - 2xy\right)^{n-2} - 1\,dx\,dy,$$

which a patient reader will reduce to the constant $2/3$. Fortunately, there is a relatively painless combinatorial proof, given below, which is both more intuitive and more powerful.

In this proof we employ the "discrete" model, in which integers between 1 and $2n$ are paired at random. Once the intervals have been selected, we label the endpoints $A(1), B(1), \ldots, A(n-2), B(n-2)$ recursively as follows:

Refer to the endpoints $\{1, \ldots, n\}$ as the *left side*, and $\{n+1, \ldots, 2n\}$ as the *right side*. Let $A(1) = n$ and let $B(1)$ be its mate. Suppose we have assigned through $A(j), B(j)$. We attach the labels $A(j+1)$ and $B(j+1)$ by the following rules:

- If $B(j)$ is on the left side: Let $A(j+1)$ be the leftmost point on the right side which has not yet been labeled. Let $B(j+1)$ be its mate.
- If $B(j)$ is on the right side: Let $A(j+1)$ be the rightmost point on the left side which has not yet been labeled. Let $B(j+1)$ be its mate.

If $A(j) < B(j)$ we say that this interval "went to the right"; otherwise, it "went to the left". Note that we are labeling endpoints of intervals from the center outwards, starting from the left when the last interval went to the right, and vice-versa. Endpoints marked $A(\cdot)$ are called *inner* points and those marked $B(\cdot)$ are called *outer*. (See FIGURE 2.)

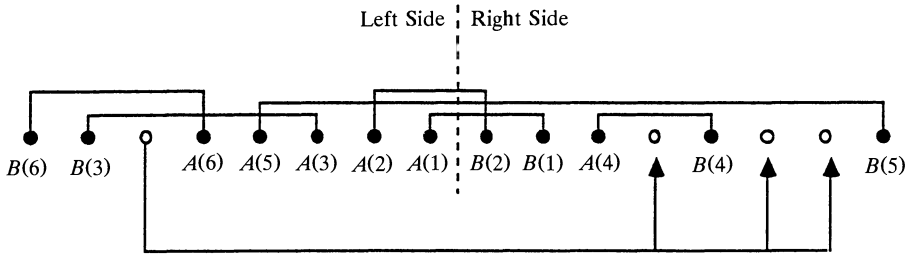


FIG. 2. Labeling intervals.

It is easy to prove by induction that immediately after the labels $A(j)$ and $B(j)$ have been assigned either:

1. an equal number of points have been assigned from the left and right sides (in case $A(j) < B(j)$), or
2. two more points have been labeled on the left than on the right (in case $A(j) > B(j)$).

Now when the labels $A(n-2)$ and $B(n-2)$ have been assigned, there remain four endpoints $a < b < c < d$ which are unlabeled. Note that the three ways of pairing them are equiprobable. Our claim is that if a is paired with either c or d, then that interval intersects all others, but if a is paired with b, then *no* interval intersects all others. This will prove the result.

We consider the only two possible cases:

1. a and b are on the left and c and d are on the right, and
2. only a is on the left.

In either case, all points labeled with $A(\cdot)$ are between a and c for otherwise one of a or c would have been labeled. It follows that either $[a, c]$ or $[a, d]$ meets all other intervals.

On the other hand, suppose $[a, b]$ and $[c, d]$ are intervals in the collection. Neither is a candidate as an interval which intersects all others since $[a, b] \cap [c, d] = \varnothing$. Suppose some interval $[e, f]$ (where $e < f$) intersects all others. Suppose e and f have received the labels $A(j)$ and $B(j)$.

In case (1), where a and b are on the left, we know that the endpoint $[e, f]$ labeled $A(j)$ is between b and c. Thus $[e, f]$ cannot intersect both $[a, b]$ and $[c, d]$.

Now consider case (2), where just a is on the left. Since $[e, f]$ meets $[c, d]$ we have $f > c$, hence f is an outer point ($f = B(j)$). Further, e is an inner point and therefore $[e, f]$ went to the right. However, the last labeled pair, $\{A(n-2), B(n-2)\}$ must have gone to the left since (in this case) we assigned more labels on the left than on the right. Thus, for some k, with $j < k \leqslant n-2$ we have that $[A(k), B(k)]$ went to the left, but $[A(k-1), B(k-1)]$ went right. Thus $A(k) < n$ and $A(k) < A(j)$ since $A(k)$ is a later-assigned, left-side inner point. It now follows that $[B(k), A(k)]$ is disjoint from $[A(j), B(j)] = [e, f]$, a contradiction.

With slightly more care one may use this construction to show that for any $k < n$, the probability that in a family of n random intervals there are at least k which intersect all others is

$$\frac{2^k}{\binom{2k+1}{k}}$$

independent, again, of n.

Acknowledgements. Edward Scheinerman is supported, in part, by ONR contract N00014-85-K-0622. Peter Winkler is supported, in part, by ONR contract N00014-85-K-0769.

REFERENCES

1. Béla Bollobás and Peter Winkler, On the longest chain among random points in Euclidean space, *Proceedings of the American Mathematical Society*, 103 (1988) 347–353.

2. P. Erdös and A. Rényi, On random graphs I, *Publ. Math. Debrecen*, 6 (1959) 290–297.
3. R. Fagin, Probabilities on finite models, *J. Symbolic Logic*, 41 (1976) 50–58.
4. Peter C. Fishburn, *Interval Orders and Interval Graphs*, Wiley, 1985.
5. Martin C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*, Academic Press, 1980.
6. J. M. Hammersley, A few seedlings of research, *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, U. of California Press, 1972, pp. 345–394.
7. B. F. Logan and L. A. Shepp, A variational problem for random Young tableaux, *Adv. in Math.*, 26 (1977) 206–222.
8. Edward R. Scheinerman, Random interval graphs, *Combinatorica*, 8 (1988) 357–371.
9. C. Schensted, Longest increasing and decreasing subsequences, *Canad. J. Math.*, 13 (1961) 179–191.
10. S. M. Ulam, Monte Carlo calculations in problems of mathematical physics, *Modern Mathematics for the Engineer*, E. F. Beckenbach, ed., McGraw Hill, New York, 1961.
11. A. M. Veršik and S. V. Kerov, Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux, *Dokl. Akad. Nauk SSSR*, 233 (1977) 1024–1028.

Playing the Numbers

You can play mathematical sequences as musical notes on your computer. If the sequence is periodic, the melody might be quite interesting, and it certainly puts a new dimension into the subject.

For example, fix a modulus m. The sequence of Fibonacci numbers

$$0, 1, 1, 2, 3, 5, \ldots \quad (\text{modulo } m)$$

is periodic. You can listen to the Fibonacci numbers (mod m) by running the following little Basic program (enter your favorite modulus m at the prompt).

```
10  CLS: INPUT "MODULUS"; M : A = 0: B = 1
20  C = A + B: C = C - M*INT(C / M)
30  SOUND 130*2 ^(C / 12), 2 : A = B: B = C: GOTO 20
```

The tune '$m = 51$,' for instance, is rather appealing. Try other periodic mathematical sequences, scoring several of them together, varying the frequencies and durations, etc. See problem E3410 on p. 916 of this issue for information about the periods.

—U. Phonious

The Dark Side of the Moebius Strip

GIDEON E. SCHWARZ, *Hebrew University, Jerusalem*

GIDEON E. SCHWARZ: Born 1933 in Salzburg, Austria. Escaped in 1938, after the Anschluss, to Palestine, today Israel. M.Sc. in Mathematics at the Hebrew University, Jerusalem in 1956. Ph.D. in Mathematical Statistics at Columbia University in 1961. Research fellowships: Miller Institute 1964–66, Institute for Advanced Studies on Mt. Scopus 1975–76. Visiting appointments: Stanford University, Tel Aviv University, University of California in Berkeley. Since 1961, Fellow of the Institute of Mathematical Statistics. Presently, Professor of Statistics at the Hebrew University.



The Moebius strip has now been around for over a century. Everybody knows how to make one, as a real-world object as well as abstractly. And yet, as a geometrical configuration in three-dimensional Euclidean space, there is still no satisfactory model for it. Furthermore, most of the attempts in this direction have remained relatively unknown. The following is an account of what I found after some search, and some research, about these matters. In particular: a smooth flat model, as had been also found by Sadowsky [1930, no. 4]; two smooth flat algebraic models, one defined by Wunderlich [1962], and one found independently by myself [1988]; some results concerning the shortest Moebius strip, by Barr [1964], who also reports results of Gardner; some by myself, first reported here; a formulation of a variational problem—the Moebius strip of least elastic energy—by Sadowsky [1930, no. 5]; and, finally, my own conjecture, as it emerges from all the facts listed here plus some observations of real-world Moebius strips.

1. What exactly is a Moebius strip? On one hand, it is often defined as the topological space attained by starting with a (closed) rectangle, endowed with the "usual" topology, and identifying two opposite edges point by point with each other, so that each vertex gets identified with the one diagonally across. This "abstract Moebius strip" serves in topology as the canonical example of a nonorientable manifold.

On the other hand, there is a physical model of the abstract strip, and it is usually denoted by the same term. Its inventor Moebius [1865] (also spelled "Möbius", but "Mobius" is not an acceptable spelling), described it as follows: a paper rectangle that is sufficiently long and narrow is bent and twisted so that its two shorter edges can be glued together in the required manner.

Between the abstract topological space and its physical model lie some concepts of intermediate degrees of abstraction. For example, besides a physical model for the topological strip, one can ask for a geometrical model in three-dimensional Euclidean space. Such a model will be provided by a subset of 3-space, that is homeomorphic to the topological strip. A simple approach to constructing such an "embedding" of the strip in 3-space consists of imitating the physical twisting that takes place when the paper strip is produced. Fix a line in three space, and pick a segment that is coplanar with the line, perpendicular to it and disjoint from it. Rotate the plane, and the segment with it, around the fixed line. At the same time, in the rotating plane, rotate the segment around its midpoint. If the second

rotation is carried out at half the angular velocity of the first one, the segment will have completed half a turn around its center when it has gone once around the line, and thus will meet itself with the required twist. If the fixed line is chosen as the z-axis of a coordinate system, the initial position of the segment is chosen to be $[R-1, R-1]$ on the x-axis, for some $R > 1$, and the angles by which the two rotations have progressed are t and $\frac{1}{2}t$, respectively, then the equations

$$x(s,t) = \left(R + s\cos\tfrac{1}{2}t\right)\cos t \qquad y(s,t) = \left(R + s\cos\tfrac{1}{2}t\right)\sin t,$$
$$z(s,t) = s\sin\tfrac{1}{2}t,$$

for $-1 \leqslant s \leqslant 1$, $0 \leqslant t \leqslant 2\pi$, define the embedding. The fact that for all s the functions at $(s, 2\pi)$ agree with their values at $(-s, 0)$ reflects the identifying of the opposite sides of the rectangle. Besides being homeomorphic to the topological Moebius strip, this model has the advantage of using only analytic functions of s and t. In fact, it is not only analytic—it is algebraic: it is possible to eliminate s and t from the expressions for the coordinates, and to represent the strip as part of an algebraic surface, that is, as the set of points where a polynomial in the coordinates takes on the value 0.

There is, however, also a drawback to this model. The nature of this drawback will become clear after taking a closer look at the physical model, the paper strip. It was produced by bending and glueing, but no stretching was used: any curve drawn on the paper rectangle would have became a space curve of the same length. A mathematical model that reflects this property of the paper strip would be an *isometric* embedding. The following considerations show that the model defined above fails on that score. The center line of the rectangle is the line $s = 0$, and its image under the embedding is the circle of radius R around the origin in the x-y plane. Its length is $2\pi R$, and an appropriate rescaling would have matched this value with the length of the rectangle in the (s,t)-plane. However, the other lines of the form $s = \textit{constant}$ go into curves of greater length. To see this, find the velocity at which a point on the strip moves when s is fixed, and t is regarded as "time." The velocity can be split into two components, one in the plane through the point and the z-axis, and one perpendicular to it. Their magnitudes are $R + s\cos\frac{1}{2}t$ and $\frac{1}{2}s$, respectively. The total velocity is, therefore, strictly greater than the first part, whenever s is not 0. Since integration of the first part from $t = 0$ to $t = 2\pi$ yields just $2\pi R$ (note that the cosine term contributes nothing to the definite integral), the space-curve corresponding to nonzero constant s is strictly longer than the circle of radius R. The model is therefore not a valid representation of any Moebius strip that can be made out of a paper rectangle. Could it fit a strip made out of a differently shaped piece of paper? Again the answer is no, but in order to see this, a local consequence of the absence of stretching will have to be considered.

A differentiable surface can be approximated in the neighborhood of any one of its points by a plane, the tangent plane at the point. If the surface is twice differentiable, it can be approximated to a higher degree of approximation by a quadric. Among twice differentiable surfaces, those that are obtained by bending, but not stretching, paper, are characterized by the property that the approximating quadrics will be planes or parabolic cylinders. Surfaces with this property, that holds if and only if the matrix of the second derivatives of z with respect to x and y is singular, are called "flat surfaces." It is one of the classical results of

differential geometry, that among twice differentiable surfaces, a piece of plane can be mapped isometrically only onto flat ones.

To the point ($s = 0$, $t = 0$) of the rectangle there corresponds the point ($x = R$, $y = 0$, $z = 0$) in space. In a neighborhood of this point, the Taylor expansion shows that $z = \frac{1}{2}(x - R)y/R$ is true modulo third and higher powers of the distance from between (x, y) and $(R, 0)$. The strip is, therefore, locally approximated by a hyperboloid at that point. Therefore, the surface is not flat, and consequently, the model fails to represent any paper model of the Moebius strip.

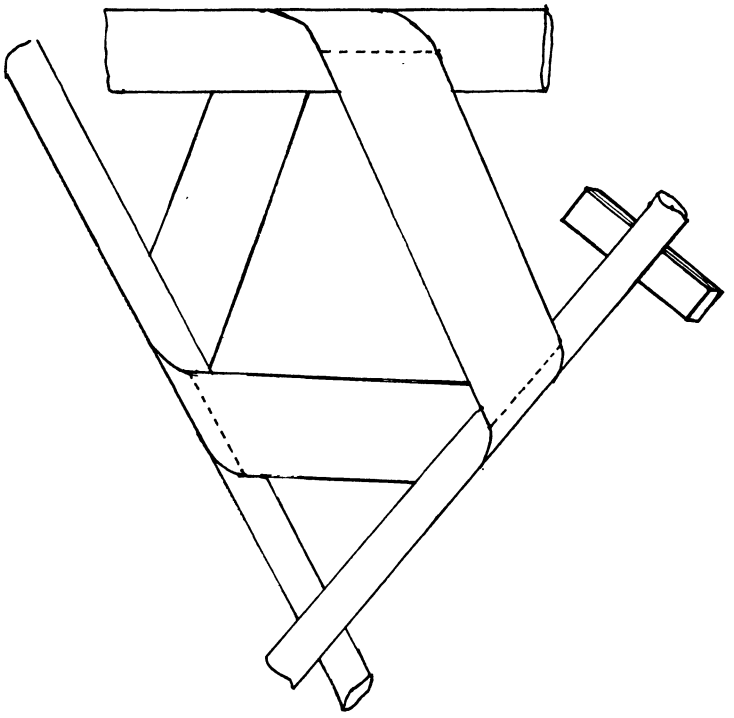
2. The simplest flat surfaces are planes and cylinders. A Moebius strip consisting of three planar parts and three cylindrical parts, smoothly joined together, was found by Sadowsky [1930, no. 1]. I had no access to his papers, and before I read about them in Wunderlich [1962], I found such a model myself. This model is best described in physical terms. Consider a long cylinder of radius r, and a long paper rectangle of width w. Imagine the rectangle lying on a table, and the cylinder lying on the rectangle at an angle, that is, so that the axis of the cylinder forms an acute angle of, say, α with the long sides of the rectangle. Clearly one can get hold of one end of the rectangle, and pulling it up, wrap the rectangle tightly half a turn around the cylinder, and then let it continue at an angle of $\pi - \alpha$ to the axis, on the plane that is parallel to the table, a distance of $2r$ above it.

Each one of the long sides of the rectangle will become a space curve consisting of a piece of a helix, and two straight-line pieces, each tangent to the helical piece at one of its end. By rolling out the cylinder onto a plane, the helical piece would become the hypothenuse of a right triangle with one side of length πr, and an angle of α facing it. Its length is therefore $\pi r/\sin \alpha$. Now project the picture down onto the table. Each of the two space curves becomes a pair of segments forming an angle of $\pi - 2\alpha$, whose missing tip is replaced by a sine curve. Note that the missing tip consists of two sides of an isosceles triangle with an angle of $\pi - 2\alpha$ facing a base of length $\pi r \cot \alpha$, and its length is $\pi r/\sin \alpha$, the same as the helical part of the space curve. Since the segment that was on the higher plane does not change its length when it is projected onto the table, and the other segment stays put, the whole space curve retains its length when it is first projected onto the table, and then its rounded tip is replaced by continuing the segments until they meet.

There is a "physical" consequence of this fact: when a (long) rectangle is folded once on the table, a cylinder, say of radius r, can be inserted into the fold, so that the rectangle wraps tightly around the cylinder for half a turn, and then continues in the horizontal plane that lies $2r$ above the table, directly over its previous position on the table.

A flat smooth Moebius strip will now be built from three of these configurations. Choose three angles α, β and γ, that sum to π; then $\pi - 2\alpha$, $\pi - 2\beta$ and $\pi - 2\gamma$ also add up to π. For the corresponding radii choose r, r and $2r$ respectively. Start with a rectangle lying on the table. Put a cylinder of radius r on it, so that its axis forms an angle of α with the long sides of the rectangle, and wrap the rectangle half a turn around it. The rectangle continues now on a plane that lies $2r$ above the table. Now repeat the procedure with the second cylinder, sufficiently far from the first one so as not to interfere with it (in "reality" a slab of thickness $2r$ will have to be inserted under the part of the rectangle that is r above the table, to keep it there). After this operation the rectangle continues on a plane

$4r$ above the table. In the third step, where a cylinder of double radius is used, wrap downwards, causing the strip to return to the plane of the table, where it can be made to meet the first part that never left the table.



The surface that was constructed here is easily seen to have a continuously varying tangent plane: it is a smooth surface. Since each half-turn wrapping flips it around once, and it meets itself after three of them, it is nonorientable, and represents the Moebius strip. On the debit side, it is not analytic at the "seams" where cylinders meet planes. In fact, the second derivatives that are needed for defining the second-order Taylor series approximation are not defined at these points; still, the surface is "flat" in the extended sense: it is locally isometric to a piece of a plane.

The construction can be modified so that higher derivatives will exist. All that is required, is that the circular cross-sections of the cylinders will be replaced by cross sections of the form

$$\frac{y^{2n}}{s^{2n}} + \frac{z^{2n}}{r^{2n}} = 0,$$

where s is chosen so that the (generalized) helix will have the right length of $\pi r/\sin \alpha$, and the surface will have $2n - 1$ derivatives. By using more complicated cross sections, derivatives of all orders may be made to exist, but this will also not yield analyticity; as long as there are planar portions of the model, it cannot be analytic.

3. Let us now have a closer look at the dimensions of the rectangle out of which the strip was made. The cylindrical parts of the surface will appear on the rectangle as three disjoint strips, two of width πr and one of width $2\pi r$, that form, with the long sides of the rectangle, angles of alternating signs, and of magnitudes α, β and γ, respectively. Let the width w of the rectangle be chosen as the unit of length. Clearly, the length of the rectangle must exceed the sum of the cotangents of α, β and γ. The average of the angles is $\pi/3$, and for positive acute angles the function cotangent is convex. By the Jensen Inequality, the average of the cotangents is at least the cotangent of the average, or $\cot(\pi/3)$, which is $1/\sqrt{3}$. The length of the rectangle therefore exceeds $\sqrt{3}$.

Conversely, starting with any rectangle whose length exceeds $\sqrt{3}$ times its width, one can, for $\alpha = \beta = \gamma = \pi/3$, and sufficiently small r, carry out the construction described above, and obtain a smooth flat Moebius strip. The question of how short a Moebius strip can be, is not completely solved by these considerations: for rectangles with *length/width* $\leqslant \sqrt{3}$ only this particular construction fails, and so a length to width ratio of more than $\sqrt{3}$ has been shown to suffice, but not to be necessary. In fact, in a certain sense there are arbitrarily short Moebius strips, if the assumption of a smooth embedding is relaxed, so as to admit "planar folded Moebius strips." These are again best described in physical terms. While the models that correspond to what can be made out of paper, without stretching and tearing, are flat, in the sense of preserving the lengths of curves, they do not have to be smooth, or even one-to-one. If the latter condition is violated by the surface cutting through itself, the model cannot represent a paper surface. Consider, however, a sheet of paper folded flat along a line. It is reasonably approximated by a mapping of a rectangle onto the plane, and this mapping is at some points two-to-one. The folded paper may then be bent, yielding a two-to-one mapping onto nonplanar surfaces. The extension to finitely many folds is obvious. Martin Gardener has constructed folded Moebius strips in which that side of the rectangle that gets identified with its opposite is arbitrarily long compared to the two other sides. The bound $\sqrt{3}$ is thus replaced by 0. After the description of Gardner's model, the bound $\sqrt{3}$ will make a come-back, in a slightly different role. Start with a rectangle of dimensions W and L. Pick a positive odd integer n, and divide the rectangle into n parallel strips of dimensions W/n by L. Folding along the dividing lines back and forth, accordion-fashion, an n-fold covered strip of length L and width W/n is obtained.

No matter how large W is, compared to L, for sufficiently large n, this strip will be narrow enough to permit its being bent and glued into a Moebius strip. Note that the fact that n is odd permits the different layers of the folded strip to be glued to each other individually, in a manner that is equivalent to the identification of sides required by the configuration of the Moebius strip.

4. Compare Gardner's folded model with another folded model, namely, the limiting configuration obtained from the three-plane-three-cylinder model, when the parameter r approaches 0. The latter, that can be constructed by folding a rectangle three times, is, by definition, the limit of a family of smooth flat models. The value of a parameter like *length/width* for such a model is the limit of its values for the flat smooth models that approximate it. Inclusion of such folded models won't change the least upper bound of a parameter. This is not the case for Gardner's model: it cannot be approximated by flat smooth surfaces. More

generally: *Any paper surface in the making of which a folding-line is bent or folded, is a finite distance away, in an appropriate sense of distance, from all smooth flat surfaces.* This follows from the fact that every smooth flat surface is a ruled surface. For every point, one of the lines that pass through the point on the paper remains a straight line as the paper is bent to make the required surface. This condition is violated when the paper is folded, and then refolded, with the second fold cutting through the first one: let the acute angle between the two fold-lines be α. Then a lines of slope β through the point of intersection of the two folding-lines will be broken at an angle of $2\max(\alpha, \beta)$, which is at least twice the angle between the folding-lines. If, in making the surface, the folding line is bent, the straight lines through any of its points will also be bent by an angle bounded from below. Thus Gardner's model does not disprove the conjecture that *the length of a smooth flat Moebius strip exceeds $\sqrt{3}$ times its width.* The model that Barr describes in Chapter 3 of his book [1964], where a paper square is folded along one diagonal, and then along the other, also does not imply the violation of this bound, since its making involves the folding of a fold. If there is a counterexample to the conjecture, it won't be found by examining flat folded models. Indeed, *the length of a flat folded Moebius strip without refolded folds is at least $\sqrt{3}$ times its width. The bound is attained by the configuration consisting of three equilateral triangles on top of each other, the first connected to the second along one side, the second to the third along another side, and the third to the first along the third side.*

This is proved by first eliminating all models in which the center-line of the rectangle becomes, after folding, a polygon of more than three sides. There remain only the models where it becomes a triangle, with the three fold-lines bisecting its three exterior angles. On the rectangle, the folding lines cut the center lines at angles α, $-\beta$ and γ. The cotangents must add up at least to the length over the width, and the proof concludes like above. Note, however, that here the bound is attained, because there is no extra room needed for any cylinders.

5. Three decades after the work of Sadowsky, the step from flat smooth models to flat analytic, and even algebraic models was accomplished by Wunderlich [1962]. Chicone [1984], who was evidently not aware of Wunderlich's article, studied the problem, and came up with a family of flat analytic surfaces in 3-space, homeomorphic to the Moebius strip, but he failed to check whether they contain parts isometric to a strip obtained from a rectangle. To understand what is missing in Chicone's model, consider the embedding of a cylindrical ("untwisted") band in the surface of a cone. It is easy to find a homeomorphic embedding, but it cannot be trimmed down to an isometric image of a cylindrical band, since there are no closed geodesics on a cone. Having seen Chicone's manuscript, but also unaware of the work of Wunderlich, I constructed a different flat algebraic model, also isometric to a rectangle [1988].

Wunderlich's model and my own are both based on the same observation: the bisecting line of the rectangle that is parallel to the long sides, becomes a (closed) geodesic on the surface. If a curve is a geodesic on a flat smooth surface, and the curve has no straight parts, then the surface is the *rectifying developable* of the curve. This is a flat surface, which forms the envelope of all the rectifying planes of the curve. The normal to the surface at any point of the curve will be the principal normal to the curve at that point. Therefore, a simple, closed curve without straight parts, whose principal normal changes its direction by 180 degrees when

the curve is traversed once, will yield the required model: its rectifying developable will contain a band around the curve, isometric to a Moebius strip made from a rectangle; and if the band is narrow enough it will not intersect itself. A simple way to make sure that the principal normal will flip around, is to make the curve symmetric under a rotation by 180° around an axis that meets the curve at two points, one where the principal normal is parallel to the axis, and one where it is perpendicular to it. The second point must be a point of zero curvature of the curve. Wunderlich chose a parametrization in which this point is approached when the parameter t of the curve approaches $\pm\infty$. The coordinates of the curve are given in his model as rational functions (that is, quotients of polynomials) of the parameter. In my model, the point of zero curvature corresponds to $t = 0$, and the fact that the curve is closed is interpreted as the curve having a periodic parametrization. This suggested trying for a parametrization by trigonometric polynomials. The simplest curve that has the required properties was found to be

$$x = \sin t, \qquad y = (1 - \cos t)^3, \qquad z = \sin t(1 - \cos t),$$

which is indeed an algebraic curve; it is the intersection of the surfaces

$$z^3 = x^3 y, \qquad \text{and} \quad 8y = x^6 + 6x^2 y + y^2.$$

As was pointed out to me by Gregory Brumfiel, the Seidenberg-Tarski Theorem [1954] can be used to show that not only the curve, but also the strip obtained in this manner is part of an algebraic surface.

6. Let us now take another step in the direction from the abstract to the real world, and include some physical properties of paper in our considerations. If the substance of paper is assumed to be elastic, with a finite value for the constant in Hook's law, it would be possible to stretch it. Paper that cannot be stretched, yet can be bent if a finite force is applied, is well approximated by an infinitely thin plate with an infinite Hook's constant, so that bending any piece of it into part of a cylinder of radius r requires an amount of energy proportional to the area of the piece divided by r^2. Since every small part of a flat surface is locally approximable by a cylinder with radius equal to the finite one among the two radii of curvature (the other one must be infinite), the total energy expended in making the surface out of paper is the integral of r^{-2} over the surface. Left to itself, the paper will take on the shape of minimum energy among all shapes in 3-space that are compatible with the metric structure of the surface.

As an example of this phenomenon, consider a cylindrical band, made out of a paper rectangle *without* twisting. The integral of $1/r$ over any closed geodesic on the band is forced to remain 2π. Under this constraint, the minimum of the integral of the square of $1/r$ is attained when r is constant, which holds when the surface is a circular cylinder. Indeed, this is the shape the surface takes on "in the real world," and if gently deformed, it returns to this shape. Paper Moebius strips also appear to have a clear stable shape, if the paper is not too soft, and the width of the strip is not too small. Sadowsky [1930, 5] approached the search for a analytical Moebius strip from this angle. He did not find an explicit form for the shape of minimum energy, but he simplified the expression for the energy: he reduced it to a one-dimensional integral, with respect to arc length s along the curve whose rectifying developable is the strip. The integrand can be expressed in

terms of the curvature $k(s)$ of the curve, its torsion $\tau(s)$ and the function

$$\psi(s) = \frac{1}{\tau\frac{dk}{ds} - k\frac{d\tau}{ds}}$$

as follows:

$$\int (k^2 + \tau^2)^2 \psi \log \frac{2k^2\psi + w}{2k^2\psi - w}\, ds,$$

where w is the width of the strip, and s ranges over its length.

Neither Sadowsky nor Wunderlich solved this variational problem—in all likelihood it is still open. On the basis of observation of physical models, Wunderlich states that "the shape appears to depend only slightly on the width w." This may be true for small values of w; in fact, Sadowsky showed that in the limit, for the "infinitely narrow band," the energy is approximately proportional to the integral of $(k^2 + \tau^2)^2/k^2$, and it is plausible that the solution to the minimum problem for finite w approaches the (also unknown) solution of this simpler problem when w goes to zero. Note, however, that for a fixed curve, when w grows, and approaches the minimum of $2k^2\psi$ along the curve, the energy becomes infinite. The dependence of the solution on w can, therefore, be expected to become much more pronounced as w becomes larger. My own observations of physical Moebius strips confirm this, and furthermore, lead to the following conjecture:

For *fixed length L, and width w less than $L/\sqrt{3}$, there exists a unique minimum-energy smooth flat Moebius strip in 3-space. As w approaches $L/\sqrt{3}$ from below, the strip approaches the configuration of three equilateral triangles on top of each other, each two connected along one side. If w exceeds $L/\sqrt{3}$, there is no smooth Moebius strip in 3-space, isometric to the corresponding rectangle.*

Acknowledgements I would like to thank Gil Bor, who made me aware of the work of Wunderlich, and through him, of Sadowsky; Gregory Brumfiel, who showed how the Seidenberg-Tarski Theorem can be used to prove that my model is part of an algebraic surface; Lester Dubins, who showed me Barr's book, where Gardner's construction is also described, and Moe Hirsch, who gave me Chicone's Manuscript.

This work was done while visiting Stanford University and the University of California in Berkeley.

REFERENCES

1. Stephen Barr, Experiments in Topology, Thomas Y. Crowell Company, New York 1964.
2. C. Chicone, Untitled manuscript, Univ. Missouri, 1984.
3. F. A. Möbius, Über die Bestimmung des Inhaltes eines Polyeders, *Ber. Verh. Ges. Wiss.* 17 (1865) 31–68. Gesammelte Werke, Bd. II, Leipzig.
4. M. Sadowsky, Ein elementarer Beweis für die Existenz eines abwickelbaren Möbiusschen Bandes und Zurückführung des geometrischen Problems auf ein Variationsproblem, *Sitzungsber. Preuss. Akad. Wiss.*, 22 (1930) 412–415.
5. _____, Theorie der elastisch biegsamen undehnbaren Bänder mit Anwendung auf das Möbiussche Band, *Verh. 3. Kongr. Techn. Mechanik*, Stockholm 1930, II, pp. 444–451.
6. G. Schwarz, A pretender to the title "Canonical Moebius Strip," *Pacific J. Math.*, 143 (1990) 195–200.
7. A. Seidenberg, A new decision method for elementary algebra, *Ann. Math.*, 60 (1954), 365–374.
8. W. Wunderlich, Über ein abwickelbares Möbiusband, *Monatshefte Math.*, 66 (1962), 276–289.

NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND RODICA SIMION

Mixing Property of the Pedal Mapping

PETER UNGAR
109 Lee Road, Scarsdale, NY 10583

In a recent article in the *Monthly* [2] Peter Lax proved a mixing property of the *pedal mapping* **P** which maps the set of angles of a triangle into the angles of the pedal triangle. His result tells us that if we start with almost any triangle, the sequence of successive pedal triangles contains triangles with shapes arbitrarily close to all possible shapes.

In this note we represent the pedal mapping as a shift. We could have stopped there and referred to books on ergodic theory for proofs of mixing properties of shift mappings but we thought it would help many readers to give brief derivations of what is needed here, using only the elements of measure theory and probability theory.

Lax follows Kingston and Synge [1] in representing a triangle of given shape as a point x in an equilateral triangle $E = \triangle PQR$. The barycentric coordinates of x are the angles of the triangle in units of π. In this representation the pedal mapping is the result of the following operations.

Connect the midpoints of the sides of E. Denote the four congruent triangles into which E is subdivided by E_P, E_Q, E_R and E_M where E_P is adjacent to P etc. On the interior of a corner triangle the pedal mapping **P** is the dilatation by a factor 2 which maps the interior of the corner triangle onto the interior of E. On the interior of E_M, **P** is a reflection in the centroid of E combined with a dilatation by 2. The boundaries of the small triangles correspond to right angled triangles and degenerate triangles for which the pedal triangle is degenerate and **P** is undefined.

We want to consider iterates of **P**. The mapping **P** is not defined on the boundaries of the four small triangles. Its square is not defined on the preimages of these, which are the boundaries of the subdivisions of $E_P, \ldots, E_M$, etc. We get rid of these loose ends by restricting **P** to the set E' obtained by removing from E its boundary points and the boundary points of all triangles obtained by successive subdivisions. The removed points form a set of measure 0.

Let us choose units so that the measure of E' is 1, so that we can employ the language of probability theory.

The mixing property we are going to prove is the following.

Let S be the intersection of E' with a triangle obtained by repeated subdivisions of E. For any point $x \in E'$ denote by $N(x, S, n)$ the number of points $x, \mathbf{P}(x), \ldots, \mathbf{P}^{n-1}(x)$ that lie in S. Then for almost all x and every S,

$$\lim_{n \to \infty} \frac{N(q, S, n)}{n} = \textit{measure of } S.$$

Once we have proved this the same statement immediately follows if S is a union of finitely many sets of the above kind. The convergence theorems on Lebesgue integrals could be used to extend the result to any measurable set S but we shall not do that; one can get that extension immediately from the argument of Lax who uses Birkhoff's ergodic theorem.

We label E_P, E_Q, E_R and E_M with the digits 0, 1, 2 and 3. Each of these triangles is subdivided into 4 smaller triangles. These will be numbered by a sequence of 2 digits. Small triangle ij is that subtriangle of i which is mapped into j by **P**. The triangles in the next level of subdivision are labeled by 3-digit base-4 numbers, etc. Thus for each point x in our probability space E' we have a base-4 sequence $x_1x_2\cdots$ which represents the nested sequence of triangles of successive subdivisions which contain x.

Let us recall what is meant by saying that a real number is *normal*. In this paragraph only, let x be a real number. The property of normality is an assertion about the digits behind the decimal point, so assume for simplicity that $0 \leqslant x < 1$. Let b be an integer > 1. Let $0.x_1x_2\ldots$ be the infinite expansion of x in base b. For any block B_k of k base-b digits, $B_k = (a_1, \ldots, a_k)$, let $N_n(x, B_k)$ be the number of occurrences of the block B_k in the sequence $x_1, \ldots, x_n$. If $N_n(x, B_k)/n \to b^{-k}$ as $n \to \infty$ for every k and every B_k then x is said to be *normal to base b*. We say x is *normal* if it is normal to every base. Borel showed that almost all real numbers are normal.

In our work x is a point in E' instead of a real number. The x_i are random variables defined on our probability space E'. Moreover, they are independent since the measure of a subtriangle specified by any given ndigit sequence is $(\frac{1}{4})^n$. It follows just as in the case of digit sequences representing real numbers that with probability 1 the digit sequence of x contains every block of length k with frequency 4^{-k}.

The digit sequence of $\mathbf{P}x$ is obtained by discarding the leftmost digit of the sequence of x. A point x is in the subdivision triangle 21301, say, if the first 5 digits of its representation are these. The image of a point under the n-th iterate of **P** is in that triangle if the five digits starting with the $n + 1$-st are as above. Borel's theorem tells us that with probability 1 the frequencies of occurrences of all possible k-digit sequences tend to $1/4^k$. This completes the proof of the mixing property of **P**.

For completeness we include a short proof that almost all numbers are normal.[†] It suffices to show that the numbers which are not normal to any base form a set of measure 0. To show this it suffices to show that for almost all numbers the frequency of any digit in the infinite representation in base b approaches $1/b$.

Indeed, let B_k be a block as in the definition of normal numbers. The base-b integer $a = a_1, \ldots, a_k$ is a possible digit in the base-b^k representation of x. If we know that for almost all x the frequency of any digit approaches b^{-k} in base b^k we can draw the following conclusion about the frequency of occurrences of B_k in the expansion to base b.

[†]For a careful and clever treatment which gives very precise information about the difference between the number of occurrences of a digit minus the expected number, see William Feller's *Introduction to the Theory of Probability*, vol. 1, ch. 6.

For almost all x the following is true: If we look only at blocks of k digits preceded by $0, k, 2k, \ldots$ other digits, the frequency of B_k in the base-b representation of x approaches b^{-k}.

Since each digit is a random variable independent of the preceding digits, a similar statement must also be true for blocks preceded by $1, k+1, \ldots$ digits, etc. So we can deduce that the frequency of any block of length k approaches the required limit for almost all x if we can prove this for any digit a to any base b.

Let X_n be the random variable which has value 1 if $x_n = a$ and is 0 otherwise. Then $X_1, X_2, \ldots$ are independent random variables with mean $1/b$ and variance $(b-1)/b^2 < 1$. The number of occurrences of a among the first n digits of x is

$$N_n = X_1 + \cdots + X_n.$$

N_n has mean n/b and variance $< n$. Chebyshev's inequality enables us to conclude

$$P\left(|N_n/n - 1/b| > n^{-1/4}\right) < n^{-1/2}. \tag{1}$$

Consider a number x such that the frequency of the digit a does not approach $1/b$ in its base-b representation. Then

$$\overline{\lim_{n\to\infty}} \left|\frac{N_n}{n} - \frac{1}{b}\right| > 0. \tag{2}$$

We observe now that (2) must also hold if the lim sup is taken over the subsequence of the integers n which are perfect fourth powers, $n = q^4$. This is because N_n/n can change between $(q-1)^4$ and q^4 by at most $4q^3/n = 4/q$ and this approaches 0. Consequently (2) implies

$$\left|\frac{N_{q^4}}{q^4} - \frac{1}{b}\right| > \frac{1}{q} \tag{3}$$

for infinitely many q. By (1) the probability that (3) holds for even a single value of $q \geqslant Q$ is $< 1/Q^2 + 1/(Q+1)^2 + \cdots$. Since this tends to 0 as $Q \to \infty$, the probability of the set of x such that the frequency of a in the representation of x does not approach $1/b$ is 0, which is what we wanted to prove.

REFERENCES

1. J. G. Kingston and J. Synge, The sequence of pedal triangles, this MONTHLY, 95 (1988) 609–622.
2. Peter D. Lax, The ergodic character of sequences of pedal triangles, this MONTHLY, 97 (1990) 377–381.

Integral Representation of a Finite Spike

R. E. MIROLLO
Department of Mathematics, Boston College, Chestnut Hill, MA 02167

S. H. STROGATZ[1]
Department of Mathematics, MIT, Cambridge, MA 02137

Consider the function f defined by the following integral:

$$f(u) = \int_{-\pi}^{\pi} \frac{(u + \cos\alpha)(1 + u\cos\alpha)}{(1 + u^2 + 2u\cos\alpha)^{3/2}}\, d\alpha, \tag{1}$$

where $u \geq 0$. We will show that

$$f(u) = \begin{cases} 2, & u = 1 \\ 0, & u \neq 1. \end{cases} \tag{2}$$

Thus the innocent-looking integral in (1) represents a discontinuous function which is zero everywhere except for a finite spike at the point $u = 1$ (FIGURE 1). We ran across this integral in the course of a stability calculation for the equilibrium states of a certain many-body dynamical system.

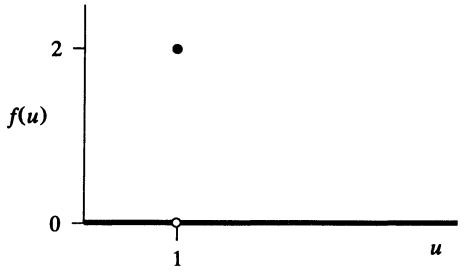


FIG. 1

The most straightforward proof of (2) involves three steps: (a) Show $f(1/u) = uf(u)$. Thus it suffices to prove (2) for $u \leq 1$. (b) Verify $f(u) = 0$ for $0 \leq u < 1$ by a series expansion in powers of u. (c) Compute $f(1) = 2$ as an elementary definite integral.

There is a more appealing proof of (2). It involves the fact that the integrand in (1) may be expressed explicitly as an exact differential. To see this, rewrite the numerator in (1) as $u + u\cos^2\alpha + u^2\cos\alpha + \cos\alpha = u\sin^2\alpha + \cos\alpha(1 + u^2 +$

[1]The research of S. H. Strogatz was supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship, DMS 8605761.

$2u\cos\alpha$), using the identity $u = u\sin^2\alpha + u\cos^2\alpha$. Thus (1) becomes

$$f(u) = \int_{-\pi}^{\pi}\left[\frac{u\sin^2\alpha}{(1+u^2+2u\cos\alpha)^{3/2}} + \frac{\cos\alpha}{(1+u^2+2u\cos\alpha)^{1/2}}\right]d\alpha$$

$$= \int_{-\pi}^{\pi}\frac{d}{d\alpha}[y_u(\alpha)]\,d\alpha$$

$$= y_u(\pi) - y_u(-\pi), \tag{3}$$

where

$$y_u(\alpha) = \frac{\sin\alpha}{(1+u^2+2u\cos\alpha)^{1/2}}. \tag{4}$$

When $u \neq 1$, $y_u(\pi) = y_u(-\pi) = 0$, and hence (3) implies that $f(u) = 0$. Now assume $u = 1$. Then $(1+u^2+2u\cos\alpha)^{1/2} = \sqrt{2}(1+\cos\alpha)^{1/2} = 2\cos(\alpha/2)$, and so $y_1(\alpha) = (\sin\alpha)/(2\cos(\alpha/2)) = \sin(\alpha/2)$ if $\alpha \neq \pm\pi$. Hence $\lim_{\alpha\to\pm\pi} y_1(\alpha) = \pm 1$. Thus (3) implies $f(1) = 1 - (-1) = 2$, as required.

There is a simple geometric interpretation of these results (Figure 2). Consider two circles of unit radius, and let u be the distance between their centers. Choose a coordinate system with the origin at the center of one of the circles, and the x-axis along the line joining the centers. Given the point $P = (u + \cos\alpha, \sin\alpha)$, define the point Q by the projection shown in FIGURE 2.

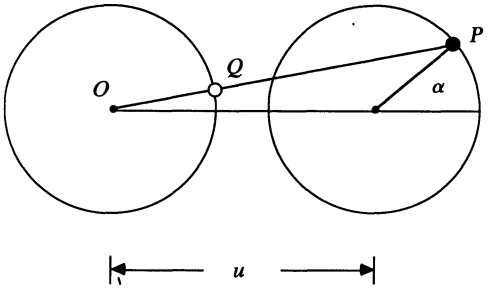


FIG. 2

Then the function $y_u(\alpha)$ may be interpreted as the y-coordinate of Q, since by similar triangles $y(Q) = y(Q)/\|OQ\| = y(P)/\|OP\| = \sin\alpha/(1+u^2+2u\cos\alpha)^{1/2} = y_u(\alpha)$. For $u \neq 1$, Q executes a closed path as α runs from $-\pi$ to π. Thus the net change in $y_u(\alpha)$ along the path is zero, which shows $f(u) = 0$ when $u \neq 1$. However, note that for u near 1, the points P, P' are splayed apart by the mapping $P \mapsto Q$ (FIGURE 3). In fact, for $u = 1$, Q is not defined when

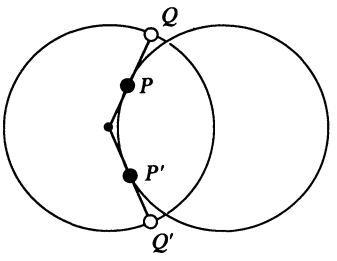


FIG. 3

$\alpha = \pm\pi$. As $\alpha \to \pm\pi$, $Q \to (0, \pm 1)$ and so $y_u(\alpha)$ changes by 2 as α runs from $-\pi$ to π. Thus $f(1) = 2$. Note that the mapping $P \mapsto Q$ cannot be extended to a continuous map at $u = 1$, $\alpha = \pm\pi$. It is this discontinuity which underlies the curious result (2).

On the Irrationality of π^2

D. DESBROW
Department of Mathematics, University of Edinburgh, Edinburgh, EH9 3JZ Scotland

1. Introduction. In 1761 Lambert [7] showed, using continued fractions, that $\tan\alpha$ is not rational when α is a nonzero rational from which it follows immediately that π is irrational. Both these results were new; the latter being but a prelude to Lindemann's celebrated proof of 1882 that π is transcendental.

That in fact π^2 is irrational by much the same reasoning with continued fractions was certainly known to Legendre [8, Note IV] by 1808. In 1873 Hermite [2] showed the irrationality of π^2 essentially from the nontrivial observation that, for the positive integer n,

$$\frac{1}{n!}\left(\frac{\pi}{2}\right)^{2n+1}\int_{-1}^{1}(1-x^2)^n\cos\frac{\pi}{2}x\,dx$$

is a polynomial in $\pi^2/4$ with integer coefficients and degree the integer part of $n/2$. It is then easy to deduce, on the assumption that π^2 is rational, the untenable existence of a positive integer less than 1. Variations on this theme are much favoured in text-books discussing the irrationality of π. Furthermore Hermite's method has been taken up and extended by others; most generally by Inkeri and Estermann. In [5] (but see also [4] for a similar, though earlier, version in English) Inkeri shows that there is no complex number α with α^2 and $\alpha^{-1}\tan\alpha$ both rational (except possibly 0 if one allows $\alpha^{-1}\tan\alpha = 1$ when $\alpha = 0$ on the grounds of continuity). Later Estermann [1], unaware of Inkeri's result, proved an equivalent form very elegantly.

The purpose here is to show yet another path through this well-trodden terrain. Specifically we offer a novel proof of the following theorem which subsumes Inkeri's result.

THEOREM. *There is no nonzero complex number* α *with* α^2 *and* $\alpha \tanh \alpha$ *both Gaussian rationals.*

By way of explanation, a complex number is a *Gaussian rational* if both its real and imaginary parts are rational.

We have the following corollaries.

(i) π^2 *(and so* π*) is irrational.*
(ii) $\tanh \alpha$ *(and so* e^α*) is not a Gaussian rational when* α *is a nonzero Gaussian rational.*

Clearly (i) follows from the choice $\alpha = \pi i$ and (ii) is immediate, noting only that

$$\tanh \frac{z}{2} = \frac{e^z - 1}{e^z + 1}$$

for each complex z not an odd integer multiple of πi.

There are of course other proofs of the irrationality of π. Two in particular are both noteworthy and ingenious. One, again by Hermite [**3**, p. 74] but less elegant than his method outlined above, uses a sequence of power series expansions. The other by Niven [**9**] uses an integral; cleverly modifying Hermite's integral method as above. Both ideas have been widely developed. Most notably in relation to our theorem and its corollaries, Inkeri [**5**] (again [**4**] should be helpful) uses Hermite's power series approach to give an alternative proof of his result cited above while Niven's integral approach has been used by Iwamoto [**6**] and Wachs [**12**] to establish the irrationality of π^2 and by Novák [**10**] to show that e^α is not a Gaussian rational when α is a non-zero Gaussian rational. Seemingly unknown to Novák, his result was already proved by Schneider [**11**] who used, however, an interpolating series for the exponential.

The proof offered here uses no more than the Taylor expansion of $\cosh \alpha\sqrt{z}$ about 1 and some ideas prompted by [**1**] and [**11**].

2. Proof of the theorem. Let α be a nonzero complex number. We begin with the observation that $w = \cosh \alpha\sqrt{z}$ *is an entire solution of*

$$4zw'' + 2w' - \alpha^2 w = 0. \tag{1}$$

[Here $\sqrt{0} = 0$ and $\sqrt{z} = \sqrt{r}e^{i\theta/2}$ when $z = re^{i\theta}$ with $r > 0$ and $-\pi < \theta \leqslant \pi$.]

For any complex z,

$$w = \cosh \alpha\sqrt{z} = \sum_0^\infty \alpha^{2n} z^n \Big/ (2n)!$$

so that w is certainly entire and the rest is routine verification.

By Taylor's theorem therefore, for some complex $a_0, a_1, a_2, \ldots$ and all complex z, real nonzero ρ,

$$w = \cosh \alpha\sqrt{z} = \sum_0^\infty a_n (z - 1)^n, \tag{2}$$

where for each n,

$$a_n = \frac{1}{2\pi i}\int_C \frac{\cosh \alpha\sqrt{z}}{(z-1)^{n+1}}\,dz \tag{3}$$

and C is the circle with centre 1 and radius ρ^2.

Substituting the expansion (2) into the equation (1) we are led immediately to the recurrence

$$\left.\begin{aligned} a_0 = \cosh\alpha, \qquad a_1 = \frac{\alpha}{2}\sinh\alpha, \\ 4(n+1)(n+2)a_{n+2} + 2(2n+1)(n+1)a_{n+1} - \alpha^2 a_n = 0. \end{aligned}\right\} \tag{4}$$

Now suppose that α^2 and $\alpha\tanh\alpha$ are both Gaussian rationals. Then $a_0 = \cosh\alpha \neq 0$ and there are positive integers k and K such that $k\alpha^2$ and $K\alpha\tanh\alpha$ are Gaussian integers. For each n let

$$b_n = n!(2k)^n K a_n/a_0.$$

Immediately from (4) we have the recurrence

$$\left.\begin{aligned} b_0 = K, \qquad b_1 = kK\alpha\tanh\alpha, \\ b_{n+2} + k(2n+1)b_{n+1} - k^2\alpha^2 b_n = 0. \end{aligned}\right\}$$

It is evident from this that, for each n, b_n and b_{n+1} are Gaussian integers not both 0, so that

$$|b_n| + |b_{n+1}| \geqslant 1. \tag{5}$$

On the other hand for each n, using (3),

$$\begin{aligned} |b_n| &= \frac{Kn!(2k)^n}{2\pi|a_0|}\left|\int_C \frac{\cosh\alpha\sqrt{z}}{(z-1)^{n+1}}\,dz\right| \\ &\leqslant \frac{Kn!(2k)^n e^{|\alpha|\sqrt{(1+\rho^2)}}}{|a_0|\rho^{2n}} \end{aligned}$$

by the usual estimates, noting that $|\cosh z| \leqslant e^{|z|}$ for each complex z.

Now for each $n \geqslant 1$ choose $\rho = n$ so that

$$|b_n| \leqslant \frac{Kn!A^n}{|a_0|n^{2n}} \leqslant \frac{K}{|a_0|}\left(\frac{A}{n}\right)^n,$$

where $A = 2ke^{2|\alpha|}$. Thus as $n \to \infty$, $b_n \to 0$; contradicting the inequality (5) and so proving the theorem.

REFERENCES

1. T. Estermann, A theorem implying the irrationality of π^2, *J. London Math. Soc.*, 41 (1966) 415–416.
2. C. Hermite, Extrait d'une lettre de Mr. Ch. Hermite à Mr. Borchardt, *J. de Crelle*, 76 (1873) 342–344; Oeuvres, t. III, Gauthier-Villars, Paris, 1912, pp. 146–149.

3. C. Hermite, Cours de M. Hermite, 4me ed., Faculté des Sciences de Paris, Hermann, Paris, 1891.
4. K. Inkeri, The irrationality of π^2, *Nordisk Mat. Tidskr.*, 8 (1960) 11–16.
5. _____, Eräiden transsendenttifunktioiden arvojen irrationaalisuudesta (Finnish), *Arkhimedes*, No. 1 (1965) 15–21.
6. Y. Iwamoto, A proof that π^2 is irrational, *J. Osaka Inst. Sci. Tech.*, (1) 1 (1949) 147–148.
7. J. H. Lambert, Mémoire sur quelques propriétés remarquables des quantités transcendentes circulaires et logarithmiques, *Histoire de l'Academie Royale des Sciences et Belles-Lettres* (Berlin) 17 (1761) 265–322.
8. A. M. Legendre, Éléments de géométrie, avec des notes, 7me ed., Didot, Paris, 1808.
9. I. Niven, A simple proof that π is irrational, *Bull. Amer. Math. Soc.*, 53 (1947) 509.
10. B. Novák, A remark to a paper of J. F. Koksma, *Nieuw. Arch. Wisk.*, 23 (1975) 195–197.
11. T. Schneider, Über der Irrationalität von π, *Bayer. Akad. Wiss. Math.-Nat. Kl. S.-B.*, 1954 (1955) 99–101.
12. S. Wachs, Contribution à l'étude de l'irrationalité de certains nombres, *Bull. Sci. Math.*, (2) 73 (1949) 77–95.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

The Width of a Rose Petal

S. C. ALTHOEN AND M. F. WYNEKEN
Department of Mathematics, University of Michigan, Flint, MI 48502

"Don't hurry, don't worry. You're only here for a short visit. So be sure to stop and smell the roses" (cf. [1]). Roses are a beautiful class of polar graphs too often passed by too quickly for lack of an interesting remark. (However, see [2] for a thorough discussion of roses and computer graphics.) We investigate three aspects of the famous cosine roses

$$r = \cos n\theta, \qquad n = 2,3,4,\ldots, \tag{1}$$

and two generalizations: the wonderful yet less well-known "double" roses

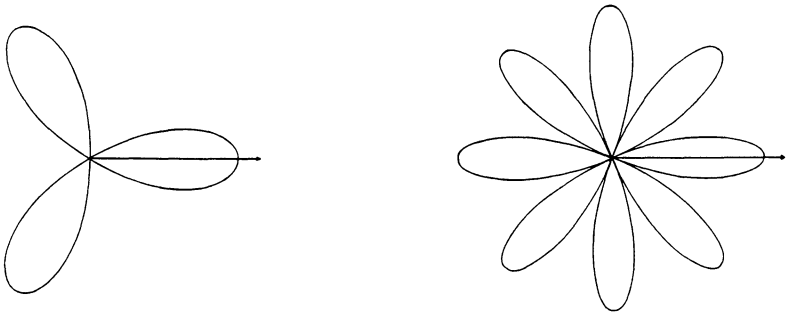
$$r = a + b\cos n\theta, \qquad n = 2,3,4,\ldots, \qquad 0 < a < b, \tag{2}$$

which generalize the limaçon, $r = a + b\cos\theta$; and the family

$$r^{\varepsilon} = \cos n\theta, \qquad n = 2,3,4,\ldots, \qquad \varepsilon > 0, \tag{3}$$

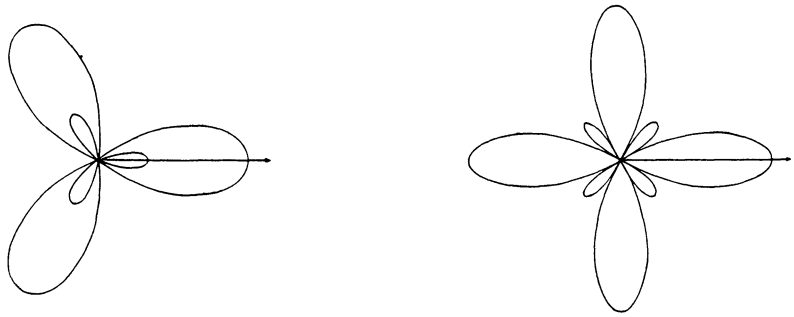
which generalizes the lemniscate, $r^2 = \cos n\theta$.

Graphs. First let's look at a few graphs:



$r = \cos 3\theta$ $\qquad$ $r = \cos 4\theta$

FIG. 1



$r = 1 + 2\cos 3\theta$ $r = 1 + 2\cos 4\theta$

FIG. 2



$r^8 = \cos 3\theta$ $r^{1/8} = \cos 3\theta$

FIG. 3

On the one hand, the little petals in FIGURE 2 grow longer as $a \to 0$. Thus $r = \cos 2k\theta$ has $4k$ petals, while $r = \cos(2k + 1)\theta$ has only $2k + 1$ petals. On the other hand, these little petals shrivel up as $a \to b$. Thus $r = 1 + \cos n\theta$ is an n-petalled rose for $n = 2, 3, \ldots$.

Petal widths. The graphs of equations (1) and (3) are symmetric about the line containing the polar axis and look generally like the graph in FIGURE 4 for $-\pi/2n \leqslant \theta \leqslant \pi/2n$.

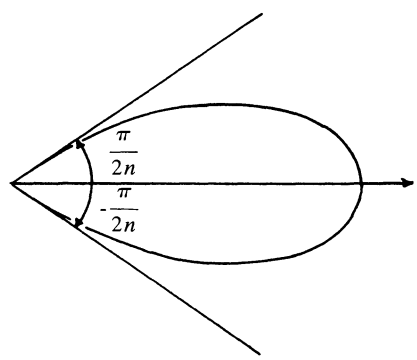


FIG. 4

A natural problem is to find the maximum width $2y_n$ of the petals of $r = \cos n\theta$. Although one can approach this problem by first converting to rectangular coordinates, it is most satisfying to attack the problem intrinsically. To this end consider FIGURE 5.

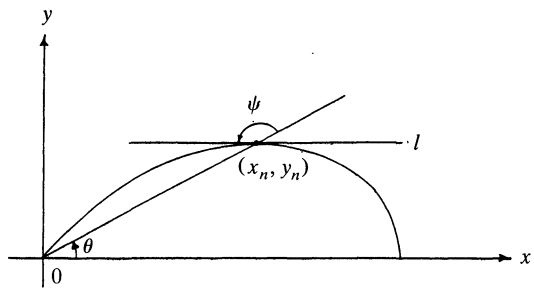


FIG. 5

The angle ψ is the angle from the radius vector to the tangent and, as is shown in calculus, $\tan\psi = r/(dr/d\theta)$. For (x_n, y_n) to be a maximum we must have the tangent line l parallel to the polar axis. Thus, we require $\psi = \pi - \theta$, so $\tan\psi = \tan(\pi - \theta) = -\tan\theta$ and so

$$\frac{r}{dr/d\theta} = -\tan\theta$$

or

$$r + \tan\theta\frac{dr}{d\theta} = 0. \tag{4}$$

In the case of the four-leaf rose $r = \cos 2\theta$, (4) yields

$$2\tan\theta = \cot 2\theta = \frac{1 - \tan^2\theta}{2\tan\theta},$$

so $\tan\theta = \sqrt{1/5}$. Then the width is $2y_n = 2r\sin\theta = 2\cos 2\theta\sin\theta = 2\sqrt{6}/9$.

In general, if $r = \cos n\theta$, (4) yields $n\tan\theta = \cot n\theta$ which is difficult to solve because the formula that expresses $\cot n\theta$ as a rational function of $\tan\theta$ is complicated. Another approach to finding the width uses the fact that $\cos n\theta$ is a polynomial in $\cos\theta$. Because $\cos n\theta$ is the real part of $e^{in\theta}$:

$$\begin{aligned}
\cos n\theta &= \operatorname{Re}(\cos\theta + i\sin\theta)^n \\
&= \operatorname{Re}\sum_{0\leq k\leq n}\binom{n}{k} i^k \sin^k\theta\cos^{n-k}\theta \\
&= \sum_{0\leq m\leq n/2}\binom{n}{2m}(-1)^m\sin^{2m}\theta\cos^{n-2m}\theta \\
&= \sum_{0\leq m\leq n/2}\binom{n}{2m}(-1)^m(1-\cos^2\theta)^m\cos^{n-2m}\theta.
\end{aligned}$$

We shall denote this polynomial by P_n, so that $P_n(\cos\theta) = \cos n\theta$. Observe that when n is even each term of P_n has even degree, and when n is odd each term of P_n has odd degree. Then, at least in principle, one can find the petal width of $r = \cos n\theta$ by doing the following. First differentiate the equation $y = r\sin\theta = \cos n\theta \sin\theta$ with respect to θ and set $dy/d\theta$ equal to zero. Denoting the smallest positive critical value by θ_n, we have

$$\cos\theta_n \cos n\theta_n = n \sin\theta_n \sin n\theta_n. \tag{5}$$

If we then add $n\cos\theta_n \cos n\theta_n$ to both sides of this equation and apply the difference formula for cosine, we obtain

$$\cos\theta_n P_n(\cos\theta_n) = \frac{n}{n+1} P_{n-1}(\cos\theta_n),$$

which is a polynomial equation in $\cos\theta_n$ of degree $n+1$ and with the degrees of terms all of the same parity. In the case when n is odd we can view this as a polynomial equation in $\cos^2\theta_n$ of degree $(n+1)/2$, and in the case when n is even we can divide both sides by $\cos\theta_n$ and then view this as a polynomial equation in $\cos^2\theta_n$ of degree $n/2$. One could, in theory, solve these equations exactly in terms of radicals for $n = 2, 3, \ldots, 8$. For arbitrary n, Newton's method could be applied to find the approximate petal width.

The reader can also do the width calculation with the equation $r = a + b\cos n\theta$, $0 < a < b$. For example, for $n = 2$ the maximum petal width of the large petal in FIGURE 6 is $(2\sqrt{6}/9)(b+a)^{3/2}b^{-1/2}$. The width of the little petal is $(2\sqrt{6}/9)(b-a)^{3/2}b^{-1/2}$. We leave the case $r^2 = a + b\cos 2\theta$, $0 < a < b$, as an exercise.

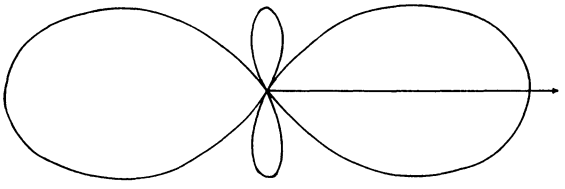


FIG. 6

Asymptotic results. In the search for asymptotic results, standard calculus limits have interesting geometric applications. First observe that $0 < n\theta_n < \pi/2$, so $\theta_n \to 0$ as $n \to \infty$. Then from (5) we obtain

$$\begin{aligned}(n\theta_n)^{-1}\cos\theta_n \cos n\theta_n &= (n\theta_n)^{-1} n \sin\theta_n \sin n\theta_n \\ &= \frac{\sin\theta_n}{\theta_n} \sin n\theta_n;\end{aligned}$$

thus $\alpha^{-1}\cos\alpha = \sin\alpha$ or

$$\alpha^{-1} = \tan\alpha, \tag{6}$$

where α is a limit point of the sequence $\{n\theta_n\}_{n\geq 1}$. From (6) we observe that α cannot equal 0 or $\pi/2$ so that α must lie in the open interval $(0, \pi/2)$, and must in fact be the limit of the entire sequence since the graphs of α^{-1} and $\tan\alpha$ intersect in only one place on this interval. The corresponding x-coordinates $x_n =$

$\cos n\theta_n \cos \theta_n$ will converge to $\cos \alpha$, and a calculator reveals $\alpha = .860333589\ldots$ and $\cos \alpha = .652184623\ldots$.

Now consider the equation $r^\varepsilon = \cos n\theta$, $\varepsilon > 0$. In solving for the smallest positive critical value $\theta_{n,\varepsilon}$ equations (5) and (6) become

$$\cos \theta_{n,\varepsilon} \cos n\theta_{n,\varepsilon} = \frac{n}{\varepsilon} \sin \theta_{n,\varepsilon} \sin n\theta_{n,\varepsilon} \tag{5$_\varepsilon$}$$

and

$$\varepsilon \alpha_\varepsilon^{-1} = \tan \alpha_\varepsilon \tag{6$_\varepsilon$}$$

where $n\theta_{n,\varepsilon} \to \alpha_\varepsilon$ as $n \to \infty$. Let $x_{n,\varepsilon}$ be the x-coordinate of the critical value $\theta_{n,\varepsilon}$, so $x_{n,\varepsilon} = (\cos n\theta_{n,\varepsilon})^{1/\varepsilon} \cos \theta_{n,\varepsilon}$, and let x_ε be the limit of $x_{n,\varepsilon}$ as $n \to \infty$, so $x_\varepsilon = (\cos \alpha_\varepsilon)^{1/\varepsilon}$. It is interesting to compute the limit of x_ε as $\varepsilon \to 0$. From (6$_\varepsilon$) one can see not only that $\alpha_\varepsilon \to 0$ as $\varepsilon \to 0$, but one can also obtain implicitly the derivative of α_ε with respect to ε. Hence, we may apply l'Hôpital's rule as follows:

$$\begin{aligned}
\lim_{\varepsilon \to 0} \ln x_\varepsilon &= \lim_{\varepsilon \to 0} \frac{\ln \cos \alpha_\varepsilon}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} -\tan \alpha_\varepsilon \frac{d\alpha_\varepsilon}{d\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{-\tan \alpha_\varepsilon}{\alpha_\varepsilon \sec^2 \alpha_\varepsilon + \tan \alpha_\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{-\sin \alpha_\varepsilon / \alpha_\varepsilon}{\sec \alpha_\varepsilon + (\sin \alpha_\varepsilon / \alpha_\varepsilon)} \\
&= -\tfrac{1}{2},
\end{aligned}$$

and so $x_\varepsilon \to 1/\sqrt{e}$ as $\varepsilon \to 0$.

We note that these calculations can also be applied to the other intersection points of α^{-1} with $\tan \alpha$. These intersection points relate to the x-coordinates of the horizontal tangents of the other petals in the limit as $n \to \infty$.

The reader can easily generalize to the equations $r^\varepsilon = a + b \cos n\theta$, $0 < a < b$, $a + b = 1$, $\varepsilon > 0$, to again obtain $\lim_{\varepsilon \to 0} x_\varepsilon = 1/\sqrt{e}$. If $a + b > 1$, the graph blows up as $\varepsilon \to 0$, and if $a + b < 1$, the graph shrinks to the pole as $\varepsilon \to 0$.

REFERENCES

1. Apologies to the late golfer Walter C. Hagen (1892–1969), *New York Times* May 22, 1977, sec. v, p. 4.
2. P. M. Maurer, A rose is a rose..., this MONTHLY, 94 (1987) 631–645.

The Sum of Absolute Deviations and the Sum of Squared Deviations

DAVID L. FARNSWORTH
Department of Mathematics, Rochester Institute of Technology, Rochester, NY 14623

This note points out a similarity in the sum-of-squares and the sum-of-absolute-values settings of data analysis. In the usual sum-of-squares setting, the sum of the squared deviations from any real number t can be written as a quadratic function

of t:

$$f(t) = \sum_{i=1}^{n} (x_i - t)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(t - \bar{x})^2 \tag{1}$$

for n data values x_i. The difference of $f(t)$ and its minimum is $n(t - \bar{x})^2$, a succinct, clearly nonnegative expression whose value is readily perceived to be zero at $t = \bar{x}$. These results, which appear in many statistics textbooks, are obtained by completing the square or by differentiating with respect to t.

The sum of the absolute values of the deviations, $g(t) = \Sigma|x_i - t|$, attains a minimum value at $t = m$, the sample median (see, e.g., [**1**, p. 209] or [**2**, p. 400]). We now present a decomposition $g(t) = \Sigma|x_i - m| + D_1(t) + D_2(t)$, which parallels (1) in that both D_1 and D_2 will be transparently nonnegative always, and equal to zero when $t = m$. Let

$$N(\tau) = \begin{cases} \#\{x_i \geqslant m\} - \#\{x_i < m\}, & \text{if } t < m \\ \#\{x_i \leqslant m\} - \#\{x_i > m\}, & \text{if } t \geqslant m, \end{cases}$$

and let Σ^* denote a sum over all x_i strictly between t and m. Then

$$g(t) = \sum_{i=1}^{n} |x_i - m| + N_t|t - m| + 2\sum\nolimits^{*}|x_i - t|. \tag{2}$$

Since $N_t \geqslant 0$, all three terms on the right are nonnegative, and since the sum Σ^* is vacuous for $t = m$, the second and third terms are zero for $t = m$, as required.

To prove (2), assume $t < m$, and note that for each x_i,

$$|x_i - t| = \begin{cases} |x_i - m| - |t - m|, & \text{if } x_i \leqslant t \\ |x_i - m| - |t - m| + 2|x_i - t|, & \text{if } t < x_i < m \\ |x_i - m| + |t - m|, & \text{if } m \leqslant x_i. \end{cases}$$

Result (2) now follows by summing over all x_i. For $t > m$ the proof is virtually identical, and if $t = m$, (2) holds trivially.

REFERENCES

1. J. D. Emerson and D. C. Hoaglin, Analysis of two-way tables by medians, in Understanding Robust and Exploratory Data Analysis, D. C. Hoaglin, F. Mosteller, and J. W. Tukey, eds., Wiley, New York, 1983.
2. R. V. Hogg and A. T. Craig, Introduction to Mathematical Statistics, 4th ed., Macmillan, New York, 1978.

The Lebesgue Integral Without Measure Theory

A. VAN DAELE
Department of Mathematics, University of Leuven, Leuven, Belgium

Very often, a course in measure theory is the first occasion where students learn about the Lebesgue integral on $\mathbb{R}$ and $\mathbb{R}^n$. However, sometimes it is convenient to have some notions of Lebesgue integration theory on $\mathbb{R}^n$ (like completeness of

$\mathscr{L}^2(\mathbb{R}^n)$ and the dominated convergence theorem) without going into measure theory. We mainly think of students in physics and engineering. There are several more or less well-known procedures to develop the Lebesgue integration theory on $\mathbb{R}^n$ without introducing too many notions of general measure theory (see, e.g., [**1**, **2**]). In most cases they start from the Riemann integral and use the notion of null sets. In this note we will show how some of the well-known techniques can be used to provide a short and fairly complete treatment of Lebesgue integration on $\mathbb{R}^n$.

We will only deal with the integral on $\mathbb{R}$. But the same method can be applied to $\mathbb{R}^n$ (and in fact to many other more general cases). For more details see [**3**].

The starting point is the Riemann integral on continuous functions.

1. DEFINITION. Denote by $K(\mathbb{R})$ the complex vector space of continuous complex functions with compact support in $\mathbb{R}$. If $f \in K(\mathbb{R})$, we use $I(f)$ for the Riemann integral of f.

Sufficient knowledge of the Riemann integral on continuous functions is assumed.

2. DEFINITION. Let $f\colon \mathbb{R} \to \mathbb{C}$ be any complex function. Define $\|f\| \in [0, \infty]$ by

$$\|f\| = \inf\left\{ \sum_{n=1}^{\infty} I(h_n) \,\middle|\, h_n \in K(\mathbb{R}),\ h_n \geqslant 0 \text{ and } |f(t)| \leqslant \sum_{n=1}^{\infty} h_n(t) \text{ for all } t \in \mathbb{R} \right\}.$$

Remark that infinite sums are allowed. There are no convergence problems because we are working in the extended positive real line. We have the following properties.

3. PROPOSITION. *Let $f, f_1, f_2, \ldots$ be complex functions on $\mathbb{R}$ and let $\lambda \in \mathbb{C}$. Then*

i. *if $f \in K(\mathbb{R})$, then $\|f\| = I(|f|)$,*

ii. $\|\lambda f\| = |\lambda|\, \|f\|$,

iii. *if $|f(t)| \leqslant \sum_{n=1}^{\infty} f_n(t)$ for all $t \in \mathbb{R}$, then $\|f\| \leqslant \sum_{n=1}^{\infty} \|f_n\|$.*

The proof of this result uses standard techniques of measure theory; see, e.g., [**1**]. Dini's theorem is used in the proof of (i) and the other properties are more or less immediate consequences of the definition. From (iii) follows that $\|f\| \leqslant \|g\|$ if $|f| \leqslant |g|$ and that $\|f + g\| \leqslant \|f\| + \|g\|$. In particular, $\|\cdot\|$ is a semi-norm.

4. DEFINITION. Denote by $\mathscr{L}^1(\mathbb{R})$ the set of complex functions f on $\mathbb{R}$ such that for all $\epsilon > 0$ there exists a $g \in K(\mathbb{R})$ such that $\|f - g\| < \epsilon$. Define $I(f) = \lim I(f_n)$ if $f \in \mathscr{L}^1(\mathbb{R})$ and if (f_n) is a sequence in $K(\mathbb{R})$ such that $\|f - f_n\| \to 0$.

It is very easy to see that $\mathscr{L}^1(\mathbb{R})$ is a complex vector space containing $K(\mathbb{R})$ as a subspace and that $\bar{f}$, $|f| \in \mathscr{L}^1(\mathbb{R})$ if $f \in \mathscr{L}^1(\mathbb{R})$. It is also easy to show that $\lim I(f_n)$ exists and does not depend on the choice of the sequence. Moreover, the definition of I on $\mathscr{L}^1(\mathbb{R})$ extends the Riemann integral on $K(\mathbb{R})$. The functional I remains linear and positive on $\mathscr{L}^1(\mathbb{R})$ and still $I(|f|) = \|f\|$ holds on $\mathscr{L}^1(\mathbb{R})$. These results follow from the corresponding results for the Riemann integral because I is the unique continuous extension of the Riemann integral to $\mathscr{L}^1(\mathbb{R})$.

We now come to the main properties.

5. THEOREM (Completeness of $\mathscr{L}^1(\mathbb{R})$). *If (f_n) is a sequence in $\mathscr{L}^1(\mathbb{R})$ such that $\|f_n - f_m\| \to 0$ as $n, m \to \infty$, then there is an $f \in \mathscr{L}^1(\mathbb{R})$ such that $\|f - f_n\| \to 0$.*

Proof. Let (f_n) be a Cauchy sequence in $\mathscr{L}^1(\mathbb{R})$. By taking a subsequence if necessary, we may assume that $\|f_{n+1} - f_n\| \leqslant 1/2^n$ for all n. Let f be any complex function on $\mathbb{R}$ such that $f(t) = \lim f_n(t)$ if $\lim f_n(t)$ exists. The function f may have any value at the other points of $\mathbb{R}$. Choose $n \in \mathbb{N}$. If $t \in \mathbb{R}$ and $\lim f_n(t)$ exists, we have

$$f(t) - f_n(t) = \sum_{k=n}^{\infty} (f_{k+1}(t) - f_k(t)) \quad \text{and}$$

$$|f(t) - f_n(t)| \leqslant \sum_{k=n}^{\infty} |f_{k+1}(t) - f_k(t)|.$$

On the other hand, if this last sum is finite, then the first one will also converge and so $\lim f_n(t)$ exists. Therefore, if $t \in \mathbb{R}$ and $\lim f_n(t)$ does not exist, then this sum is infinite and the inequality is trivial. So our inequality is true for all $t \in \mathbb{R}$. By (iii) of the proposition we get

$$\|f - f_n\| \leqslant \sum_{k=n}^{\infty} \|f_{k+1} - f_k\| \leqslant \frac{1}{2^{n-1}}.$$

From this it follows that $f \in \mathscr{L}^1(\mathbb{R})$ and that $f_n \to f$ in $\mathscr{L}^1(\mathbb{R})$.

By similar methods we can prove the monotone convergence theorem.

6. THEOREM. *Suppose that (f_n) is a sequence in $\mathscr{L}^1(\mathbb{R})$ such that $0 \leqslant f_1 \leqslant f_2 \leqslant \cdots$ and that $\sup I(f_n) < \infty$. If f is a function on $\mathbb{R}$ such that $f(t) = \lim f_n(t)$ for all t, then $f \in \mathscr{L}^1(\mathbb{R})$ and $I(f) = \lim I(f_n)$.*

Proof. For all n and all $t \in \mathbb{R}$ we have

$$f(t) - f_n(t) = \sum_{k=n}^{\infty} (f_{k+1}(t) - f_k(t))$$

and by the proposition

$$\begin{aligned} \|f - f_n\| &\leqslant \sum_{k=n}^{\infty} \|f_{k+1} - f_k\| \\ &= \sum_{k=n}^{\infty} I(f_{k+1} - f_k) = \lim_m I(f_m) - I(f_n). \end{aligned}$$

The right-hand side converges to 0 if $n \to \infty$. It follows that $f \in \mathscr{L}^1(\mathbb{R})$. So also $\|f - f_n\| = I(f - f_n) = I(f) - I(f_n)$ and we obtain $I(f) = \lim I(f_m)$.

By a standard procedure the dominated convergence theorem can be obtained from the monotone convergence theorem (see, e.g., [**1**, **2**]).

7. THEOREM. *Suppose that (f_n) is a sequence in $\mathscr{L}^1(\mathbb{R})$ and that f is a complex function such that $f(t) = \lim f_n(t)$ for all $t \in \mathbb{R}$. Suppose that $g \in \mathscr{L}^1(\mathbb{R})$ and that $|f_n(t)| \leqslant g(t)$ for all $t \in \mathbb{R}$. Then $f \in \mathscr{L}^1(\mathbb{R})$ and $I(f) = \lim I(f_n)$.*

Proof. By considering real and imaginary parts we may assume that the functions $f, f_1, f_2, \ldots$ are real. If we apply the monotone convergence theorem to the sequence $(g + \sup_{n \leqslant k \leqslant l} f_k)_{l=n}^{\infty}$ we obtain that $\sup_{n \leqslant k} f_k \in \mathscr{L}^1(\mathbb{R})$. If we apply the monotone convergence theorem once more to the sequence $(g - \sup_{n \leqslant k} f_k)$

we find that $\limsup f_n \in \mathscr{L}^1(\mathbb{R})$ and so $f \in \mathscr{L}^1(\mathbb{R})$. Moreover we obtain $I(f_k) \leqslant I(\sup_{n \leqslant k} f_k)$ if $n \leqslant k$ so that

$$\limsup I(f_n) = \inf_n \left(\sup_{k \geqslant n} I(f_k) \right) \leqslant \inf_n \left(I\left(\sup_{k \geqslant n} f_k \right) \right) = I(f).$$

Similarly, or by applying the same formula to $-f_n$ and $-f$, we get $I(f) \leqslant \liminf I(f_n)$. Therefore, $I(f) = \lim I(f_n)$.

We only used techniques that are well known in integration theory. Nevertheless, using some simple new ideas (as in the proof of the completeness of $\mathscr{L}^1(\mathbb{R})$), we have obtained a shorter and simpler treatment of the main results on the Lebesgue integral on $\mathbb{R}$. We finish this note by showing that also $\mathscr{L}^2(\mathbb{R})$ can be treated in such a short way. In fact, the same is true for all $\mathscr{L}^p(\mathbb{R})$ with $1 \leqslant p < \infty$ (see [3]).

8. DEFINITION. Let $\|f\|_2 = \|f^2\|^{1/2}$ for all $f: \mathbb{R} \to \mathbb{C}$.

In the following proposition we have collected all the necessary results on $\| \cdot \|_2$ to obtain $\mathscr{L}^2(\mathbb{R})$ and its properties.

9. PROPOSITION. *Let $f, g, f_1, f_2, \ldots$ be complex functions on $\mathbb{R}$ and $\lambda \in \mathbb{C}$.*

i. *If $\|f\|_2 < \infty$ and $\|g\|_2 < \infty$ then $\|fg\| \leqslant \|f\|_2 \|g\|_2$,*

ii. $\|\lambda f\|_2 = |\lambda| \, \|f\|_2$,

iii. *if $|f(t)| \leqslant \sum_{n=1}^{\infty} |f_n(t)|$ for all t, then $\|f\|_2 \leqslant \sum_{n=1}^{\infty} \|f_n\|_2$.*

Proof. If $\|f\|_2 \leqslant 1$ and $\|g\|_2 \leqslant 1$ we find

$$\|fg\| = \| \, |f| \, |g| \, \| \leqslant \tfrac{1}{2} \| \, |f|^2 + |g|^2 \| \leqslant \tfrac{1}{2} \left(\| \, |f|^2 \| + \| \, |g|^2 \| \right) \leqslant 1.$$

In the general case we just have to multiply f and g by the appropriate constants to prove (i). Again (ii) is trivial. To prove (iii), observe that

$$\|f^2\| \leqslant \sum_{n,m} \| \, |f_n| |f_m| \, \| \leqslant \sum_{n,m} \|f_n\|_2 \|f_m\|_2 = \left(\sum_n \|f_n\|_2 \right)^2$$

if all $\|f_n\| < \infty$. Hence (iii) follows in all cases.

It is clear that now we can proceed by defining $\mathscr{L}^2(\mathbb{R})$ as the closure of $K(\mathbb{R})$ with respect to this new seminorm. And by the same arguments, $\mathscr{L}^2(\mathbb{R})$ will be a complex vector space and $\mathscr{L}^2(\mathbb{R})$ will also be complete. Moreover, the product of two functions in $\mathscr{L}^2(\mathbb{R})$ will be in $\mathscr{L}^1(\mathbb{R})$ because of the definition and the Cauchy-Schwarz-like inequality proved in the proposition.

REFERENCES

1. E. W. Packel, Functional Analysis, A Short Course, International Textbook Company, New York, 1974.
2. A. J. Weir, Lebesgue Integration and Measure, Cambridge University Press, Cambridge, 1973.
3. A. Van Daele, The Lebesgue Integral on $\mathbb{R}$ and $\mathbb{R}^n$, Lecture Notes K. U. Leuven, unpublished, 1989.

PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)

COLLABORATING EDITORS: I. DAVID BERG, RICHARD L. BISHOP, DUANE M. BROLINE, FRANK S. CATER, GULBANK D. CHAKERIAN, EVERETT C. DADE, UNDERWOOD DUDLEY, IRA M. GESSEL, RICHARD A. GIBBS, CLARK GIVENS, DOUGLAS A. HENSLEY, JOHN R. ISBELL, DANIEL J. KLEITMAN, FRED KOCHMAN, JOSEPH D. E. KONHAUSER, R. G. LANGEBARTEL, FREDERICK W. LUTTMAN, MARVIN MARCUS, FRANK B. MILES, RICHARD PFIEFER, STEPHEN L. PORTNOY, BRUCE A. REZNICK, J. J. ROTMAN, J. O. SHALLIT, LAJOS TAKACS, DANIEL ULLMAN, AND EDWARD T. H. WANG.

Proposed problems should be sent to the MONTHLY PROBLEMS *address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

*An asterisk (*) after the number of a problem indicates that neither the proposer nor the editors supplied a solution.*

Solutions of published problems should arrive before May 31, 1991 at the MONTHLY PROBLEMS *address given on the inside front cover. Two copies suffice. Please type with double line spacing and include the solver's name and mailing address on each sheet. Include a self-addressed postcard or label if an acknowledgement is desired.*

A publishable solution must, above all, be correct and complete. Given these attributes, elegance and conciseness are preferred. The answer to the problem should appear at the beginning of the solution. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, please tell the editors and include a copy of the solution if feasible.*

ELEMENTARY PROBLEMS

E 3409. *Proposed by Ioan Tomescu, University of Bucharest, Bucharest, Romania.*

Suppose G is a connected k-chromatic graph which is neither a complete graph nor a cycle on m vertices with $m \equiv 3 \pmod 6$. Prove that in any k-coloring of G there exist two vertices of the same color having a common neighbor.

E 3410. *Proposed by Peter Freyd, University of Pennsylvania, Philadelphia, PA.*

(a) Given a positive integer m, let $f(m)$ be the period-length of the Fibonacci sequence taken modulo m. Prove that $f(m) \leq 6m$ for all m and that equality holds for infinitely many values of m.

(b) Prove the analogous assertion for the Lucas sequence with 6 replaced by 4. (The Fibonacci sequence $\{F_n\}_{n=0}^{\infty}$ satisfies $F_0 = 0$, $F_1 = 1$, and $F_{n+1} = F_n + F_{n-1}$ for $n \geq 1$; the Lucas sequence $\{L_n\}_{n=0}^{\infty}$ satisfies $L_0 = 2$, $L_1 = 1$, $L_{n+1} = L_n + L_{n-1}$ for $n \geq 1$.)

E 3411. *Proposed by Donald E. Knuth and Boris Pittel, Stanford University, Stanford, CA.*

Find a closed formula for

$$\sum_{k_1+k_2+\cdots+k_n=m} \frac{1}{k_1!(k_1!+k_2!)\cdots(k_1!+k_2!+\cdots+k_n!)}$$

and

$$\sum_{k_1+k_2+\cdots+k_n=m} \frac{1}{2^{k_1}(2^{k_1}+2^{k_2})\cdots(2^{k_1}+2^{k_2}+\cdots+2^{k_n})},$$

where both sums are extended over all n-tuples of nonnegative integers with sum m.

E 3412. *Proposed by Eric Wepsic, Boston, MA.*

Suppose that we place n balls (numbered from 1 to n) into n urns (numbered from 1 to n) in the following way: The ith ball is placed into an urn chosen randomly from the first i urns. Let $P(n, k)$ be the probability that an urn chosen at random from the n urns contains exactly k balls (i.e., $nP(n, k)$ is the expected number of urns containing exactly k balls). Find $\lim_{n\to\infty} P(n, k)$ for fixed k.

E 3413. *Proposed by Robert McNaughton, Rensselaer Polytechnic Institute, Troy, NY.*

Suppose n_1, n_2 are positive integers and m_1, m_2 are integers such that $m_1/n_1 < m_2/n_2$. Let V be the semigroup under componentwise addition formed by all pairs (p, q), where p and q are integers, $q > 0$, and $m_1/n_1 \le p/q \le m_2/n_2$. Call a set B a generating set for V if $B \subseteq V$ and every element of V is equal to a finite sum of elements of B, repetitions allowed.

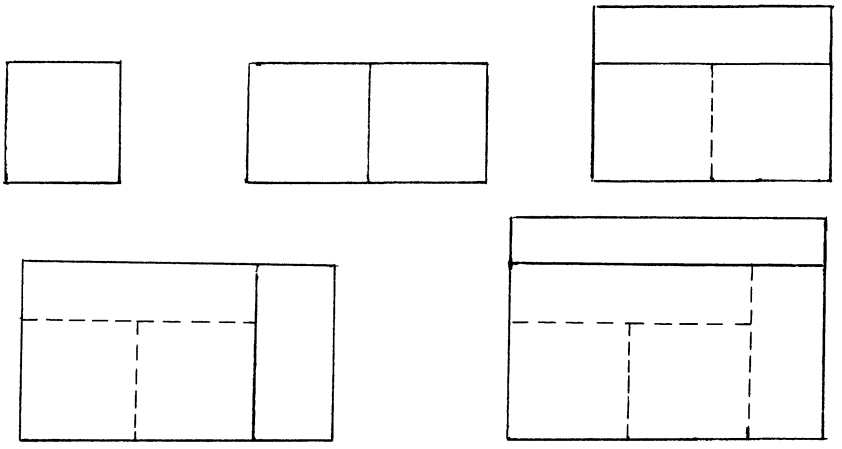
(a) Prove that V has a generating set B in which $q \le \max(n_1, n_2)$ for every $(p, q) \in B$.

(b) Prove that V has a unique minimal generating set, i.e., one contained in every other generating set.

E 3414. *Proposed by Gerald Myerson, Macquarie University, New South Wales, Australia.*

Suppose we construct a sequence of rectangles as follows. We begin with a square of area one. We then alternate adjoining a rectangle of area one alongside or on top of the previous rectangle. The figure shows the first five rectangles in the sequence.

Find the limiting ratio of length to height.



SOLUTIONS OF ELEMENTARY PROBLEMS

Looking at the Regular Polyhedra

E 3167 [1986, 650]. *Proposed by Ernesto Bruno Cossi, Universidade Federal do Rio Grande do Sul, Brazil.*

Let P_i be any one of the five regular polyhedra inscribed in a unit sphere. For each polyhedron P_i determine the smallest and the largest number of vertices of P_i which can be seen from a point on a concentric sphere of radius $R > 1$.

Solution by Matthew Cook (student), University of Illinois at Urbana-Champaign, and the editors. Presumably the problem asks for the complete evaluation of the two functions f_i and F_i on $(1, \infty)$ defined as follows: $f_i(R)$ is the minimal number of vertices of P_i visible from a point on a concentric sphere of radius R, while $F_i(R)$ is the maximal number of vertices of P_i visible from a point on a concentric sphere of radius R. (The unit sphere is considered to be transparent; only the surface of P_i has "hiding power.") Clearly both f_i and F_i are non-decreasing step functions taking only a finite number of values.

Suppose P_1 is the regular tetrahedron, P_2 is the cube, P_3 is the regular octahedron, P_4 is the regular dodecahedron, and P_5 is the regular icosahedron. For simplicity let us refer to the center of the unit sphere as the *origin*.

The first two cases are comparatively trivial. For the regular tetrahedron $f_1(R) = 3$ and $F_1(R) = 4$ for all R, while for the cube $f_2(R) = 4$ and $F_2(R) = 7$ for all R. In both cases the minimum number $f_i(R)$ occurs on a ray from the origin through the centroid of a face, while the maximum number $F_i(R)$ occurs on a ray from the origin through a vertex.

The planes of the faces of each regular polyhedron divide the exterior of the polyhedron up into a finite number of parts. In the interior of any such part the number of vertices visible remains constant. One could go beyond the question raised in the problem and ask for the complete description of all of these parts and the number of vertices visible in the interior of each one. This description is rather easy in the case of the tetrahedron and cube.

In the case of the octahedron we could take the vertices to be the points $(\pm 1, 0, 0)$, $(0, \pm 1, 0)$, and $(0, 0, \pm 1)$. Then the planes of the eight faces have the equations $x \pm y \pm z = \pm 1$. In the interiors of the various parts into which these planes divide the exterior of the octahedron, either three, four, five or six vertices are visible.

Using this information it is easy to see that $f_3(R)$ and $F_3(R)$ each take on two distinct values. Thus $f_3(R) = 3$ for $1 < R < \sqrt{3}$ (equality occurring on a ray from the origin through the centroid of a face) and $f_3(R) = 4$ for $\sqrt{3} < R < \infty$ (equality occurring on a ray from the origin through the midpoint of an edge). Similarly $F_3(R) = 5$ for $1 < R < \sqrt{3}$ (equality occurring on a ray from the origin through a vertex) and $F_3(R) = 6$ for $\sqrt{3} < R < \infty$ (equality occurring on a ray from the origin through the centroid of a face).

Now we turn to the dodecahedron, for which the situation is somewhat more complicated. Although it is hard to visualize the regions of space formed by extending a dodecahedron's faces, it is not hard to categorize them. But first, it will be very useful to observe that if three faces A, B, C are arranged so that A is

adjacent to B, and B is adjacent to C, but A is not adjacent to C, then whenever all three are visible, we must have at least one of two situations: Either the face adjacent to A, B, and C must be visible or the two other faces adjacent to B must be visible. (See FIGURE 1.)

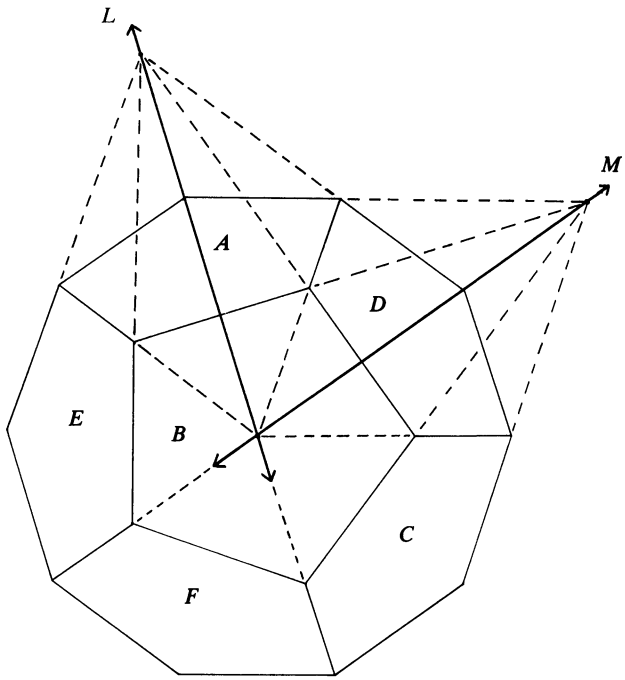


FIG. 1

In terms of the picture, we claim that when D is not visible (but A, B, and C are) we can see E and F. We shall only show that E must be visible, since a symmetric argument would do the same for F.

Let P be the point of intersection of the extensions of faces A, E, F, C, and D, directly out from the center of B. Let L be the line of intersection of the extensions of E and D, and let M be the line of intersection of the extensions of A and C. Clearly, L and M both pass through P.

The region in which neither E nor D is visible is the V-shaped trough formed by two half planes bounded by L, one containing E and the other containing D. The region in which both A and C are visible is the V-shaped trough formed by two half planes bounded by M, one's extension containing A and the other's extension containing C. These troughs touch only at P. Hence there is no region in which A and C are visible but D and E are not. This completes the proof of our claim.

Now we can start categorizing the regions of space formed by the extensions of the faces of the dodecahedron, with the criterion of number of visible faces. There is only one kind of region in which just a single face can be seen; it is a five-sided pyramid extending from the surface of the face to the point of intersection of the extensions of the five adjacent faces. (P was such a point.) Since the regions are

defined by which faces are visible within them, each different kind of region will be defined by the pattern of visible faces. Clearly there is only one kind of region in which exactly two faces can be seen, since they must be adjacent faces. Such a region consists of the space between two nearby pyramids of the kind described above.

There is only one kind of region in which exactly three faces are visible, since of the two possible contiguous arrangements of three faces, one necessitates the visibility of even more faces, as was shown in the claim above.

If four faces are visible, there must be two which are adjacent not to each other but to a third visible face. This is the premise of our initial claim, so if we are to see only one face besides these three, we know which it must be. Thus there is only one kind of region in which four faces are visible.

If exactly five faces are visible, then we must be able to see one face from each of five of the six pairs of opposite faces. Let us denote the hidden face of the diagram above by $A', B', \ldots,$ so that A' is opposite A, B' is opposite B, and so on. Now suppose that D and D' are the opposite faces neither of which we see. Of the other ten faces, five are adjacent to D, and five are adjacent to D'. We must be seeing at least three contiguous faces from one of these two groups. Without loss of generality, suppose we see faces A, B, C. Then, by our claim, since we do not see D, we just see E and F. Thus there is only one possible arrangement for five visible faces.

Finally, for six faces, there are two different kinds of regions corresponding to the only two possible arrangements of six visible faces. Clearly, no more than six faces can be seen at once from any viewpoint: at most one face from each of the six pairs of opposite faces can be seen.

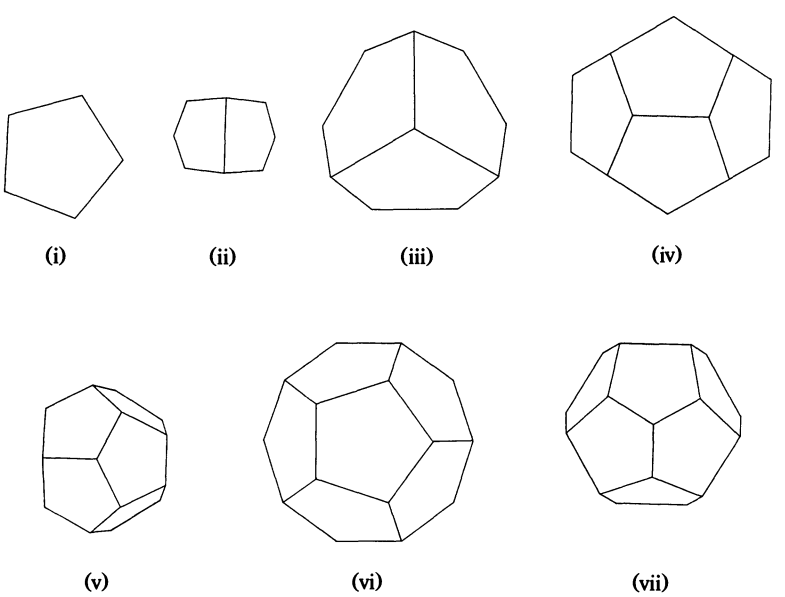


FIG. 2

Number of visible faces:	1	2	3	4	5	6	6
Arrangement of visible faces: from FIGURE 2	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Number of visible vertices:	5	8	10	12	14	15	16
Minimum distance from origin:	1	1	1	1.512	1.777	1.777	4.236
Maximum distance from origin:	1.777	1.512	4.236	∞	∞	∞	∞

Compiling this data, we see that

$$f_4(R) = 5 \quad \text{for } 1 < R < 1.777$$
$$f_4(R) = 10 \quad \text{for } 1.777 < R < 4.236$$
$$f_4(R) = 12 \quad \text{for } 4.236 < R < \infty$$

and that

$$F_4(R) = 10 \quad \text{for } 1 < R < 1.512$$
$$F_4(R) = 12 \quad \text{for } 1.512 < R < 1.777$$
$$F_4(R) = 15 \quad \text{for } 1.777 < R < 4.236$$
$$F_4(R) = 16 \quad \text{for } 4.236 < R < \infty.$$

The case of the icosahedron is still more complicated and we shall not present it here. Visual inspection of physical models and a modicum of reasoning gives us

$$3 \leqslant f_5(R) \leqslant 8, \qquad 6 \leqslant F_5(R) \leqslant 11.$$

Thus fully half of the twelve vertices can be seen somewhere on the concentric sphere of radius $1 + \varepsilon$, and all but one of the vertices can be seen from a suitable vantage point on the concentric sphere of radius 10, say.

No solutions were received before the customary deadline. The above solution was prepared by Mr. Cook at the request of the editors.

A Discrete Maximum Problem

E 3253 [1988, 258]. *Proposed by Sydney Bulman-Fleming and Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Canada.*

(a) Let n and t be natural numbers with n odd and $t \leqslant n$. Suppose x_i is an integer not divisible by n for $i = 1, 2, \ldots, t$. Determine the maximum possible number of pairs (i, j) with $1 \leqslant i < j \leqslant t$ such that $x_i + x_j \equiv 0 \pmod{n}$.

(b) Let k, n, and t be natural numbers with n relatively prime to k and $t \leqslant n$. Suppose x_i is an integer not divisible by n for $i = 1, 2, \ldots, t$. Determine the maximum possible number of k-tuples $(i_1, i_2, \ldots, i_k)$ with $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant t$ such that $x_{i_1} + x_{i_2} + \cdots + x_{i_k} \equiv 0 \pmod{n}$.

Solution of (a) by Walther Janous, Ursulinengymnasium, Innsbruck, Austria. The answer is $\lfloor t^2/4 \rfloor$. Without loss of generality we may assume $1 \leqslant x_i \leqslant n - 1$. Put $n = 2r + 1$, where $r > 0$. (The problem is vacuous when $n = 1$.) Let A be the number of x_i such that $1 \leqslant x_i \leqslant r$. Since $x_i + x_j \equiv 0 \pmod{n}$ if and only if

$x_j = 2r + 1 - x_i$, the maximum number of admissible pairs (i, j) for a given value of A is $A(t - A)$. If t is even, we have $A(t - A) \leqslant t^2/4$, with equality, for example, when $t/2$ of the x_i have the value 1 and the $t/2$ remaining x_j have the value $n - 1$. If t is odd, we have $A(t - A) \leqslant \{(t - 1)/2\}\{(t + 1)/2\}$, with equality, for example, when $(t - 1)/2$ of the x_i have the value 1 and the $(t + 1)/2$ remaining x_j have the value $n - 1$. In either case the maximum number of admissible pairs (i, j) is $\lfloor t^2/4 \rfloor$.

The assumption $t \leqslant n$ is unnecessary, but the assumption that n be odd is essential. For if n is even, we could take all the x_i to be $n/2$ and obtain $t(t - 1)/2$ admissible pairs.

Editorial remark on Part (b). When $n > k > 2$ the maximum number of admissible k-tuples $(i_1, i_2, \ldots, i_k)$ appears to occur (at least approximately) when $\lfloor (k - 1)t/k \rfloor$ of the x_i have the value 1 and the $t - \lfloor (k - 1)t/k \rfloor$ remaining x_i have the value $n - k + 1$. In that case there are

$$\binom{\lfloor (k-1)t/k \rfloor}{k-1}\{t - \lfloor (k-1)t/k \rfloor\} \sim \left(\frac{k-1}{k}\right)^{k-1} \frac{t^k}{k!}$$

admissible k-tuples. Again the assumption $t \leqslant n$ seems unnecessary, but the assumption $(n, k) = 1$ appears essential. For if $(n, k) = d > 1$, we could take all the x_i to have the value n/d and obtain

$$\binom{t}{k} \sim \frac{t^k}{k!}$$

admissible k-tuples.

Part (a) was solved also by the proposers. No solutions were received for part (b), which should have been marked with an asterisk.

A Sequence Generated by Translation and Reversal

E 3254 [1988, 258]. *Proposed by David Singmaster, Polytechnic of the South Bank, London, England.*

If a_0 and c are given positive integers, define a sequence $\{a_n\}$ by letting a_{n+1} be the positive integer obtained by reversing the digits in the ordinary decimal expansion of $a_n + c$ $(n = 0, 1, 2, \ldots)$. For example, if $a_0 = 18$ and $c = 2$, the sequence begins 18, 2, 4, 6, 8, 1, 3, 5, 7, 9, 11, 31, 33, 53, 55, 75, 77, 97, ...

(a) If $c = 1$, show that the sequence $\{a_n\}$ is eventually periodic and that all such sequences have the same ultimate cycle.

(b) If $c = 2$ and $1 \leqslant a_0 \leqslant 99$, show that the sequence $\{a_n\}$ is eventually periodic.

(c)* If $c = 2$, show that the sequence $\{a_n\}$ is eventually periodic for any choice of a_0.

Solution to (a) found by all respondents. Let a have $k \geqslant 2$ digits. The successor of a has no more than k digits, and if a ends in 9, then its successor has fewer than k digits. Since iteration brings a 9 to the rightmost position (in at most 17 steps), the process ultimately reaches the cycle $(1, \ldots, 9)$ independent of a_0.

Solution to (b) by J. C. Binz, University of Bern, Switzerland and Walther Janous, Ursulinengymnasium, Innsbruck, Austria (independently). For $c = 2$ and $1 \leqslant a_0 \leqslant 99$

Because $b'0 < 10^s$ and $\gamma\delta < 5 \cdot 10^s$, we have $\gamma\delta + b'0 < 6 \cdot 10^s$, which implies that $\gamma\delta + b'0 = ef$, where $e \leqslant 5$ and f has s digits. In particular, $d'(c + \gamma)\delta + b'0 = d'(c + e)f$, with $c + e \leqslant 9$. Furthermore, $(e + c)d + 2 < 10 \cdot 10^s$, so $(e + c)d + 2 = (e + \gamma)\delta$, with $e + \gamma \leqslant 9$ a single digit. We now have

$$a_{n+2} = (f'(e + c)d + 2)' = (f'(e + \gamma)\delta)' = \delta'(\gamma + e)f = \delta'(\gamma + \gamma)\delta + b'0,$$

as desired.

This leaves the case that $cd = 5 \cdot 10^s - 1 = 49 \cdots 9$ ($s + 1$ digits). Now $n < 5 \cdot 10^{2s-1} - 1$ implies $b < 10^{s-1} - 1$, and thus $n + 2 = (b + 1) \cdot 5 \cdot 10^s + 1$. With γ, δ playing the roles of c, d for $n + 2$ as above, we have $\gamma = 0$ and $\delta = 0 \cdots 01$ (s digits). The two-step iteration is

$$a_{n+2} = ((a_n + 2)' + 2)' = ((d'8d + 2 + b'0)' + 2)' = ((d'9\delta + b'0)' + 2)'.$$

Because δ has as many digits as $b'0$, we have $9\delta + b'0 = 9b'1$, and

$$a_{n+2} = (1b9d + 2)' = (1(b + 1)0\delta)' = \delta'0\delta + (b + 1)'0,$$

as desired.

Editorial comment. Bondesen also pointed out that, although every starting value proceeds to a cycle when $c = 2$, there are arbitrarily large cycles; in fact, the value $10^{2s+1} + 10^s + 1$ occurs in a cycle of length $9 \cdot 10^s$ for $s = 1, 2, 3, \ldots$, so that in particular 1011 occurs in a cycle of length 90. Further L. L. Foster noted that $a_0 = 1011$ is the smallest value of a_0 not ultimately leading to the cycle of length eighty-one given in the above solution of (b).

The proposer noted that when $c = 3$, already the starting values $a_0 = 1$ and $a_0 = 2$ lead to cycles of different lengths.

Complete solutions were received from A. Bondesen, L. L. Foster, and A. Riese. Parts (a) and (b) only were solved by J. C. Binz, W. Janous, R. S. Lehman, S. Pedersen (Denmark) and C. H. Webster.

The Lattice of Connected Subgraphs of a Graph

E 3281 [1988, 655]. *Proposed by Sandi Klavžar and Marko Petkovšek, University of Ljubljana, Ljubljana, Yugoslavia.*

Given a graph G, let $C(G)$ be the inclusion ordering on the vertex subsets of G inducing connected subgraphs. Prove that $C(G)$ is a lattice if and only if G is connected and every cycle of G induces a complete subgraph of G.

Solution by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA. Suppose $C(G)$ is a lattice. Note that a single vertex trivially induces a connected subgraph. If a and b are any two vertices of G, then their supremum in $C(G)$ induces a connected subgraph containing both a and b. Thus all vertices of G belong to the same component, and G is connected. If a and b are distinct vertices contained in a cycle of G, then there are two sets of vertices S_1 and S_2 that induce connected subgraphs and intersect in $\{a, b\}$. The infimum of two sets in $C(G)$ is contained in their intersection. Since each of a and b is a lower bound for S_1 and S_2 and $C(G)$ is a lattice, the infimum must equal $\{a, b\}$. Hence $\{a, b\}$ induces a connected subgraph, which implies that ab is an edge and the cycle induces a complete subgraph.

Conversely, suppose that G is connected and every cycle of G induces a complete subgraph. We first establish the existence of infima. Let $\{S_i\}$ be any nonempty collection of vertex subsets of G that induce connected subgraphs. If $\bigcap S_i$ contains just one vertex or no vertices, then this is the infimum. Otherwise, suppose $\{a, b\} \subseteq \bigcap S_i$. Fixing a particular S_j, let P be a minimal subset of vertices that induces a connected subgraph such that $\{a, b\} \subseteq P \subseteq S_j$; this set must induce a chordless path from a to b.

We claim that $P \subseteq \bigcap S_i$. Otherwise, there is some k such that $P \not\subset S_k$, but then S_k also contains a path Q from a to b. Let u be the vertex of P before its first vertex not in Q, and let v be the next vertex of P that also belongs to Q. Then u and v belong to a cycle, and the hypothesis on G implies that uv is an edge. This chord of P contradicts its minimality, thus establishing the claim.

Since $P \subseteq \bigcap S_i$, we conclude that a and b are in the same component of the subgraph induced by $\bigcap S_i$. With a, b chosen arbitrarily, this subgraph is connected. It follows that $\bigcap S_i$ is the infimum of $\{S_i\}$. Finally, let $\{T_r\}$ be the collection of all common upper bounds for $\{S_i\}$. Since the set of all vertices is such an upper bound, $\{T_r\}$ is nonempty. The supremum of $\{S_i\}$ is the infimum of $\{T_r\}$, and we conclude that G is a lattice.

Editorial comment. The graphs in which every cycle induces a clique are also known as "Husimi trees." For general graphs, the ordering $C(G)$ has been studied in other contexts.

Solved also by R. J. Hendel, A. Pedersen (Denmark), P. M. Shimura (student), and the authors.

The Angle Bisectors of a Triangle

E 3282 [1988, 762]. *Proposed by D. M. Milošević, Pranjani, Yugoslavia.*

Given a triangle ABC, suppose that w_a, w_b, w_c are the lengths of the angle bisectors, s is the semiperimeter, r is the radius of the incircle, and R is the radius of the circumcircle. Prove that

$$w_a^2 + w_b^2 + w_c^2 \leqslant s^2 - r(R/2 - r),$$

with equality if and only if the triangle is equilateral.

Solution by Walther Janous, Ursulinengymnasium, Innsbruck, Austria. Starting from Stewart's formula for the length of an angle bisector

$$w_a^2 = bc - a^2bc/(b + c)^2 = bc - a^2bc/(2s - a)^2,$$

we get

$$w_a^2 + w_b^2 + w_c^2 = bc + ca + ab - abc\{f(a) + f(b) + f(c)\},$$

where $f(t) = t/(2s - t)^2$ for $0 < t < 2s$. Since $f''(t) = (8s + 2t)/(2s - t)^4$, the function f is strictly convex on the interval $(0, 2s)$. Thus

$$f(a) + f(b) + f(c) \geqslant 3f(\{a + b + c\}/3) = 3f(2s/3) = 9/(8s)$$

and

$$w_a^2 + w_b^2 + w_c^2 \leqslant bc + ca + ab - 9/(8s),$$

with equality if and only if the triangle is equilateral. Now equating two well-known

formulas for the area of a triangle gives $s^2r^2 = s(s-a)(s-b)(s-c)$ or

$$sr^2 = (s-a)(s-b)(s-c) = s^3 - (a+b+c)s^2 + (bc+ca+ab)s - abc$$
$$= -s^3 + (bc+ca+ab)s - abc$$

or $bc + ca + ab = r^2 + s^2 + abc/s$. Thus

$$w_a^2 + w_b^2 + w_c^2 \leqslant r^2 + s^2 - abc/(8s).$$

Using the familiar relation $abc/s = 4rR$, we get the claimed inequality.

Editorial comment. Stewart's formula can easily be derived using the law of sines and the law of cosines. The proposer remarked that the above inequality and Euler's inequality $2r \leqslant R$ immediately give $w_a^2 + w_b^2 + w_c^2 \leqslant s^2$, with equality only if the triangle is equilateral. (Cf. 5.1 and 5.34 of O. Bottema et al., *Geometric Inequalities*, Groningen, 1969.) H. W. Guggenheimer mentioned that the inequality $w_a^2 + w_b^2 + w_c^2 \leqslant s^2$ was part of E1628 [1963, 891; 1964, 687].

Solved also by J. Anglesio (France), R. Bellot-Rosado (Spain), E. Braune (Austria), J. Fukuta (Japan), L. Kuipers (Switzerland), I. Paasche (W. Germany), R. Prielipp, I. A. Sakmar (Turkey), V. Schindler (E. Germany), J. M. Stark, M. Vowe (Switzerland), and the proposer. Three incorrect solutions were received.

A Double Impossibility

E 3331 [1989, 524]. *Proposed by Walter Rudin, University of Wisconsin, Madison.*

Exhibit a positive continuously differentiable function f on $[0, \infty)$ such that

$$f'(x) \geqslant f^2(x) \qquad (0 \leqslant x < \infty),$$

or prove that no such function exists.

Solution by Eugene Levine, Adelphi University, Garden City, NY. No such function exists, since if f is such a function we obtain the contradiction

$$\frac{1}{f(0)} = \int_0^{1/f(0)} dx \leqslant \int_0^{1/f(0)} \frac{f'(x)}{f^2(x)}\, dx = \frac{1}{f(0)} - \frac{1}{f(1/f(0))} < \frac{1}{f(0)}.$$

Editorial comment. Several solvers proved more general results. One possible extension is that, given $\alpha > 0$, $\varepsilon > 0$, and $m \geqslant 0$, no positive differentiable function exists satisfying $f'(x) \geqslant \alpha f^{1+\varepsilon}(x)$ on $[m, \infty)$. Also, several solvers pointed out that the problem has the same answer if $f^2(x)$ is understood to mean $f(f(x))$ instead of $(f(x))^2$.

Solution by William S. Zwicker, Union College, Schenectady, N.Y. under the alternative interpretation. Assuming that $f^2(x)$ means $f(f(x))$, we prove that there is no positive differentiable function f on $[0, \infty)$ which satisfies $f'(x) \geqslant f(f(x))$ for $0 \leqslant x < \infty$. Assume f is such a function. Since $f'(x) \geqslant f(f(x)) > 0$, f is strictly increasing, so that $f'(x) \geqslant f(f(x)) \geqslant f(f(0)) > 0$. This means that $f'(x)$ is bounded away from 0, and so $\lim_{x\to\infty} f(x) = \infty$. Hence $\lim_{x\to\infty} f(f(x)) = \infty$, and so $\lim_{x\to\infty} f'(x) = \infty$. If we set $h(x) = f(x) - (x+1)$, then $h'(x)$ also approaches ∞, which implies that $h(x)$ approaches ∞. Thus we may choose an x with $f(x) > x + 1$. Now applying the mean value theorem to the interval $[x, f(x)]$, we obtain a point c in the interval $(x, f(x))$ for which $f(f(x)) = f(x) + f'(c)(f(x) - x) > f'(c)(f(x) - x) \geqslant f(f(c))(f(x) - x) > f(f(x))(f(x) - x) > f(f(x))$, which is a contradiction.

Solved also by 105 others and the proposer. Most solvers (including the proposer) interpreted the problem as Levine did, but over 10% pointed out the ambiguity.

A Deranged Recurrence

E 3335 [1989, 525]. *Proposed by D. E. Knuth, Stanford University, Stanford, CA.*

Solve the recurrence

$$x_0 = a,\ x_1 = b,\ x_{n+2} = x_{n+1} + x_n/(n+1) \text{ for } n \geqslant 0$$

both exactly (in terms of familiar functions of n) and asymptotically.

Solution I by Jungmin Woo (student), Yonsei University, Seoul, South Korea. The solution is $x_n = (n+1)[a + (b-2a)\Sigma_{k=0}^{n+1}(-1)^k/k!] \sim (n+1)[a + (b-2a)/e]$. Let $y_n = x_n/(n+1)$ and $z_n = y_n - y_{n-1}$ for $n > 0$, $z_0 = a$. Then $y_0 = a$, $y_1 = b/2$, $z_1 = b/2 - a$, and $(n+1)y_n = ny_{n-1} + y_{n-2}$ for $n \geqslant 2$. This yields $(n+1)z_n = -(y_{n-1} - y_{n-2}) = -z_{n-1}$, so

$$z_n = \left(-\frac{1}{n+1}\right)\left(-\frac{1}{n}\right)\cdots\left(-\frac{1}{3}\right)z_1 = (-1)^{n+1}\frac{(b-2a)}{(n+1)!}$$

for $n \geqslant 1$. Since $y_n = y_0\Sigma_{i=1}^{n} z_i$, we have $y_n = a + (b-2a)\Sigma_{i=1}^{n}(-1)^{i+1}/(i+1)!$. This yields x_n as claimed above, since $\Sigma_{i=1}^{n}(-1)^{i+1}/(i+1)! = \Sigma_{k=0}^{n+1}(-1)^k/k!$.

Solution II by Hans Kappus, Rodersdorf, Switzerland. This is a second order homogeneous linear recurrence, which means that $x_n = Ay_n + Bz_n$ for linearly independent solutions $\{y_n\}, \{z_n\}$ and suitable constants A, B. With $y_n = n+1$ serving as one solution, put $z_n = w_n/n!$ to obtain the recurrence $w_{n+2} = (n+2)(w_n + w_{n+1})$, the numbers $u_n = w_{n-1}$ thus satisfying the well-known recurrence $D_{n+2} = (n+1)(D_n + D_{n+1})$ for the derangement numbers $D_n = n!\Sigma_{k=0}^{n}(-1)^k/k!$. Hence $z_n = D_{n+1}/n!$ satisfies the given recurrence. Using the initial conditions, we obtain $x_n = a(n+1) + (b-2a)D_{n+1}/n!$. Since $D_n = n!/e + O(1/n)$, we have $x_n = (n+1)[a + (b-2a)/e] + O(1/(n+1)!)$.

Solution III by J. M. Borwein and P. B. Borwein, Dalhousie University, Halifax, Nova Scotia. The generating function $A(y) = \Sigma x_n y^{n+1}/(n+1)$ satisfies the differential equation $A' - a - by = y(A' - a) + yA$ obtained by multiplying the recurrence by y^{n+2} and summing over $n \geqslant 0$. The initial condition is $A(0) = 0$. By standard techniques for solving first-order differential equations, we obtain $A(y) = (a-b) + a/(1-y) + (b-2a)e^{-y}/(1-y)$. For $n \geqslant 0$, the coefficient of y^{n+1} in the power series expansion of A is $a + (b-2a)\Sigma_{k=0}^{n+1}(-1)^k/k!$, which yields the same formula for x_n as above. Asymptotics can be determined as above.

Solved by the proposer and seventy-seven readers. One incorrect solution was received.

A Triangle from a Tetrahedron

E 3337 [1989, 641]. *Proposed by Murray Klamkin, University of Alberta, Edmonton, Alberta.*

Suppose the two longest edges of a tetrahedron are a pair of opposite edges. Prove that the three edges incident to some vertex of the tetrahedron are congruent to the sides of an acute triangle.

Solution by Jesús Ferrer, Universidad Complutense, Madrid, Spain. Let O, A, B, C be the vertices of the tetrahedron, and suppose that OA and BC are the longest edges. If the edges incident to each vertex fail to be the side lengths of an acute triangle, then the law of cosines implies the following inequalities:

$$\begin{aligned} OA^2 &\geqslant OB^2 + OC^2, \qquad BC^2 \geqslant OB^2 + AB^2, \\ OA^2 &\geqslant AB^2 + AC^2, \qquad BC^2 \geqslant OC^2 + AC^2. \end{aligned} \tag{$*$}$$

(The inequalities $(*)$ are valid *a fortiori* if some triples do not satisfy the triangle inequality.) Summing these inequalities, we obtain

$$OA^2 + BC^2 \geqslant OB^2 + OC^2 + AB^2 + AC^2.$$

Expressing the vector $\overrightarrow{BC}$ as $\overrightarrow{OC} - \overrightarrow{OB}$ (and similarly for $\overrightarrow{AB}$ and $\overrightarrow{AC}$), we can use the fact that squared length equals vector inner product to rewrite this as

$$\begin{aligned} 0 &\geqslant OA^2 + OB^2 + OC^2 - 2\overrightarrow{OA} \cdot \overrightarrow{OB} - 2\overrightarrow{OA} \cdot \overrightarrow{OC} + 2\overrightarrow{OB} \cdot \overrightarrow{OC} \\ &= \left\|\overrightarrow{OA} - \overrightarrow{OB} - \overrightarrow{OC}\right\|^2, \end{aligned}$$

that is, $\overrightarrow{OA} - \overrightarrow{OB} - \overrightarrow{OC} = 0$. Thus $\overrightarrow{OC} = \overrightarrow{BA}$ and $\overrightarrow{OB} = \overrightarrow{CA}$, which requires $OBAC$ to be a plane parallelogram. Since equality must hold in all four parts of $(*)$, in fact $OBAC$ is a rectangle.

In all other cases at least one of the inequalities $(*)$ fails, and the lengths of segments incident to the corresponding vertex are the side lengths of an acute triangle.

Solved also by C. Athanasiadis (student), J. Balogh (student, Hungary), L. Bankoff, R. J. Chapman (United Kingdom), O. P. Lossers (The Netherlands), J. H. Nieto (Venezuela), A. Nijenhuis, and the proposer. Two incorrect solutions were received.

POSTMORTEM COMMENTS ON ELEMENTARY PROBLEMS

Weitzenböck Generalized

E 3150 [1986, 400; 1988, 658]. *Proposed by George A. Tsintsifas, Thessaloniki, Greece.*

Let ABC be a triangle with sides a, b, c and area F. It is well known that $a^2 + b^2 + c^2 \geq 4F\sqrt{3}$. If p, q, r are arbitrary positive real numbers, prove that

$$\frac{p}{q+r}a^2 + \frac{q}{r+p}b^2 + \frac{r}{p+q}c^2 \geq 2F\sqrt{3}.$$

Remark by the editors. The first of the two published solutions was submitted by Victoria Pambuccian and was correctly attributed to her in the copy submitted by the editors and in the galley proofs. Unfortunately, someone subsequently changed the name of the solver to Victor Pambuccian without consulting the editors.

3-Player Blackjack

E 3186 [1987, 71; 1989, 736]. *Proposed by I. A. Sakmar, University of Western Ontario.*

On a current TV game show, three contestants for the final showcase prizes compete by spinning a wheel, which is divided into twenty equal sections num-

bered in steps of 5 from 5 to 100. The aim is to get as close to 100 as possible without exceeding a sum of 100 in two spins. Each contestant completes one or both spins before the next contestant's turn, and if either the 1st or 2nd contestant exceeds 100, that contestant is out of contention. In case of a tie, the first contestant to attain that score is chosen.

In the running of the program, it is implied that it is a good policy for the first contestant to stop if the number from the first spin is greater than 50. What is the number at which the first contestant should stop on her first spin?

Remark by the editors. Peter A. Griffin has pointed out that (a) the published solution ignores the possibility of collusion (an unlikely event on the actual show) and (b) the numerical value given for the probability that A wins (assuming no collusion) should be 0.341223 instead of 0.341771.

Producing a Ranking From Recent Results

E 3240 [1987, 996; 1989, 529]. *Proposed by C. L. Mallows, AT&T Bell Laboratories, Murray Hill, NJ.*

The tennis players in a certain club compete in singles matches. After each match is played, a ranking of the players is computed according to the following rule. Starting with the most recent match and working back through time, use the match results to build up a partial order. Ignore any match that is inconsistent with more recent results. Stop when the partial order is a complete order.

Prove or disprove: a player cannot improve his position by intentionally losing a match.

Remark by the editors. The proposer's solution published in 1989 gave a scenario in which a player benefited by throwing a match in a competition of seven players. In an editorial comment, we asked if there was such an example with fewer than seven players. B. Hayes has provided the following example with only six players. Suppose that the following results are obtained in the order given, where ab means a beats b:

$$61\ 26\ 64\ 36\ 34\ 42\ 23\ 31\ 21\ 52\ 15\ 41.$$

The ranking after the last match is $3 > 6 > 4 > 1 > 5 > 2$, but if 4 now plays 5 and loses, the ranking becomes $5 > 4 > 2 > 3 > 6 > 1$. Player 4, by losing, rises from third place to second.

ADVANCED PROBLEMS

6643. *Proposed by Gérard Letac, Université Paul Sabatier, Toulouse, France.*

For positive real t put

$$\phi(t) = \frac{1}{\Gamma(t)} \int_t^{+\infty} x^{t-1} e^{-x}\, dx.$$

Prove that ϕ increases from 0 to $\frac{1}{2}$ when t varies from 0 to $+\infty$.

6644. *Proposed by Peter Rogerson, SUNY at Buffalo.*

A needle with length L between 1 and $\sqrt{2}$ is tossed at random upon the Cartesian plane. Find the probability that it comes to rest not crossing any line of

the form $x = m$ or $y = n$, where m and n are integers. (The case $L < 1$ is treated on pp. 255–256 of J. V. Uspensky's *Introduction to Mathematical Probability*, McGraw Hill, 1937. Uspensky attributes that case of the problem to Laplace.)

6645. *Proposed by Robert Kreczner, University of Wisconsin-Stevens Point.*

(a) Show that the sequence $\{(\sin n)^n\}_{n=1}^{\infty}$ is dense in $(0, 1)$.
(b)* Show that the same sequence is dense in $(-1, 0)$.

SOLUTIONS OF ADVANCED PROBLEMS

Great Expectations

6576 [1986, 1036]. *Proposed by Hans U. Gerber, University of Lausanne, Switzerland.*

Suppose $X_1, X_2, \ldots$ are independent identically distributed real random variables with $E(X_k) = \mu$. Put $S_k = X_1 + X_2 + \cdots + X_k$ for $k = 1, 2, \ldots$

(a) If $\rho < \mu < 1$, where $\rho = -0.278465\ldots$ is the real root of $xe^{1-x} = -1$, show that the series

$$\sum_{k=1}^{\infty} S_k^k e^{-S_k}/k!$$

converges with probability one.

(b) If $X_1, X_2, \ldots$ are positive and if $\mu < 1$, show that the expectation of

$$\sum_{k=1}^{\infty} S_k^k e^{-S_k}/k!$$

is $\mu/(1 - \mu)$.

*(c) In (b) is it possible to relax the condition that the random variables are positive? For example, would it suffice to assume $E(|X_k|) < \infty$ and $\rho < \mu < 1$?

Solution I by Christian Hipp, Institut für Mathematische Stochastik, Hamburg, Germany. For part (a) note that $\rho < \mu < 1$ implies $|\mu e^{1-\mu}| < 1$. So by the strong law of large numbers,

$$\limsup_k \left| \frac{1}{k} S_k \exp\left(1 - \frac{1}{k} S_k\right) \right| < 1$$

with probability 1. Stirling's inequality

$$k! \geqslant (2\pi)^{1/2} k^{k+1/2} e^{-k}$$

implies that for $k \geqslant 1$,

$$\left| S_k^k \frac{1}{k!} e^{-S_k} \right| \leqslant \left| \frac{1}{k} S_k \exp\left(1 - \frac{1}{k} S_k\right) \right|^k.$$

Hence,

$$\sum_k \left| S_k \frac{1}{k!} \exp(-S_k) \right| < \infty$$

with probability 1.

For part (b) the random variables $X_1, X_2, \ldots$ are positive. Let $N(t)$, $t \geqslant 0$, be a Poisson process with intensity 1. Then, a straightforward calculation shows that, given $X_1, X_2, \ldots$,

$$\frac{1}{k!} S_k^k e^{-S_k} = \mathbb{P}\{N(X_1 + \cdots + X_k) = k\}.$$

Let $Y_1, Y_2, \ldots$ be the interarrival times of $N(t)$, $t \geqslant 0$. The Y's are i.i.d. and have an exponential distribution with mean 1. Since

$$N(X_1 + \cdots + X_k) = k \Leftrightarrow Y_1 + \cdots + Y_k \leqslant X_1 + \cdots + X_k < Y_1 + \cdots + Y_{k+1},$$

the mean M of

$$\sum_{k=1}^{\infty} \frac{1}{k!} S_k^k e^{-S_k}$$

equals the expected number of times $k \geqslant 1$ for which

$$Y_1 + \cdots + Y_k \leqslant X_1 + \cdots + X_k < Y_1 + \cdots + Y_{k+1}.$$

Now consider the compound Poisson process $S(t)$, $t \geqslant 0$, with interarrival times $Y_1, Y_2, \ldots$ and jumps $X_1, X_2, \ldots$. In this process, M is the mean number of crossings of zero of the path

$$t - S(t), \qquad t \geqslant 0$$

or, equivalently, the mean number of ruins of $t - S(t)$ (ruin occurs whenever $t < S(t)$). By the strong Markov property, the number of ruins has a geometric distribution with parameter

$$\mathbb{P}\{\exists t : t < S(t)\} = EX = \mu.$$

Hence, $M = \mu/(1 - \mu)$.

Solution II (for (b)) by O. P. Lossers, Eindhoven University of Technology, Eindhoven, the Netherlands. Let φ be defined by $\varphi(s) = Ee^{-(s+1)X}$. Then $\varphi^k(s) = Ee^{-(s+1)S_k}$, and so

$$M := E \sum_{k=1}^{\infty} S_k^k e^{-S_k}/k! = \sum_{k=1}^{\infty} \int_0^{\infty} \frac{x^k e^{-x}}{k!}\, dF^{*k}(x),$$

where F^{*k} is the distribution function of S_k. It follows that

$$M = \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \left[\left(\frac{d}{ds} \right)^k \varphi^k(s) \right]_{s=0}.$$

By Lagrange's formula (see e.g., G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Erster Band, Springer, 1964, problem 209, p. 125) we have

$$1 + \sum_{k=1}^{\infty} \frac{u^k}{k!} \left[\left(\frac{d}{ds} \right)^k \varphi^k(s) \right]_{s=0} = \frac{1}{1 - u\varphi'(t)},$$

where u and t are related by $u\varphi(t) = t$. Since $\varphi(-1) = 1$ and $-\varphi'(-1) = \mu$, putting $u = t = -1$ yields $M = \mu/(1 - \mu)$ as required, for $0 \leqslant \mu < 1$.

Editorial comment. No solutions were received for part (c), but O. P. Lossers considered as an example the case where X_i has a negative exponential distribution with mean $\mu > 0$. He gave M for $\mu > 1$. In fact, the same calculation holds if

$-X_i$ has a negative exponential distribution, with $EX_i = \mu < 0$, where μ lies in the interval specified below.

In particular, if X_i (or $-X_i$) has a negative exponential distribution with mean μ, then S_k (or $-S_k$) has a Gamma distribution with parameters k and μ. Hence, letting $I = (0, \infty)$, (or $(-\infty, 0)$),

$$ES_k e^{-S_k} = \int_I s^k e^{-s} \frac{s^{k-1}}{\Gamma(k)\mu^k} e^{-s/\mu}\, ds$$

$$= \frac{\Gamma(2k)}{\Gamma(k)} \left(\frac{\mu}{(1+\mu)^2}\right)^k.$$

Now,

$$\frac{\Gamma(2k)}{\Gamma(k)k!} = \frac{(2k-1)!}{(k-1)!k!} = \frac{1}{2}\frac{(2k)!}{k!k!} = \frac{1}{2}\binom{-1/2}{k}(-4)^k.$$

Therefore,

$$M = \sum_{k=1}^{\infty} \frac{1}{k!}\frac{\Gamma(2k)}{\Gamma(k)}\left(\frac{\mu}{(1+\mu)^2}\right)^k = \frac{1}{2}\sum_{k=1}^{\infty}\binom{-1/2}{k}\left(\frac{-4\mu}{(1+\mu)^2}\right)^k$$

$$= \frac{1}{2}\left\{\left(1 - \frac{4\mu}{(1+\mu)^2}\right)^{-1/2} - 1\right\} = \frac{1}{2}\left\{\left|\frac{1+\mu}{1-\mu}\right| - 1\right\},$$

where the series converges if (and only if) $-\frac{1}{4} \leqslant \mu/(1+\mu)^2 < \frac{1}{4}$. This inequality holds except when $-3 - \sqrt{8} < \mu < -3 + \sqrt{8}$ (approximately $-5.83 < \mu < -.17$) or when $\mu = 1$. Thus

$$M = \begin{cases} \dfrac{\mu}{1+\mu} & -3 + \sqrt{8} \leqslant \mu < 1 \\ \dfrac{1}{1-\mu} & 1 < \mu < \infty \text{ or } -\infty < \mu \leqslant -3 + \sqrt{8} \\ \infty & \mu = 1 \\ \text{undefined} & -3 - \sqrt{8} < \mu < -3 + \sqrt{8}. \end{cases}$$

O. P. Lossers conjectures that this is always true whenever the original series converges.

Parts (a) and (b) were solved also, via upcrossing probabilities, by Boris Pittel and by the proposer. Solutions via Laplace transforms and Lagrange's formula were given also by Terence R. Shore, Western Maryland College Problems Group, and a referee.

Rays in Normed Linear Spaces

6586 [1988, 963]. *Proposed by Charles S. Allen, Drury College, Springfield, MO.*

In an infinite-dimensional normed linear space does there exist a convex set which does not contain a ray but whose closure does contain a ray?

Solution by Thomas W. Starbird, Jet Propulsion Laboratory, Pasadena, CA. We show the answer is yes by following Ann Griesel's suggestion to construct such a

set in an arbitrary infinite-dimensional topological linear space X. In fact, we shall construct in such an X a sequence x_{ij} $(i \geqslant 1,\ j \geqslant 1)$ whose convex hull contains no ray, but whose closure (and hence the closure of its convex hull) contains a ray.

Let $\{e_i\}$, $i = 0, 1, 2, \ldots$, be a collection of linearly independent points in X. Let $\{r_i\}$, $i = 1, 2, \ldots$ be an enumeration of the positive rationals. Set

$$x_{ij} = r_i e_0 + e_i/j, \qquad i \geqslant 1,\ j \geqslant 1.$$

Since $x_{ij} \to r_i e_0$ as $j \to \infty$, the closure of $\{x_{ij}\}$ contains the ray from 0 in the direction e_0. We now show that the convex hull of $\{x_{ij}\}$ cannot contain a ray.

Fix two distinct points a and b in a ray R, and assume that they are in the convex hull of $\{x_{ij}\}$. Hence they are convex combinations of finitely many $\{x_{ij}\}$, and belong to the linear span L of $\{e_0, e_1, \ldots, e_n\}$ for some integer n. Thus R itself is contained in L. If R is also in the convex hull of $\{x_{ij}\}$ then any z in R has the form

$$z = \sum_i \sum_j \alpha_{ij} x_{ij},$$

where the α_{ij} are nonnegative and

$$\sum_i \sum_j \alpha_{ij} = 1.$$

But the α_{ij} are 0 for $i > n$ since the e_i are linearly independent—the formula for x_{ij} has only nonnegative coefficients so no cancellation can occur. Thus R is in the convex hull of the x_{ij} with $i \leqslant n$. But this is merely the convex hull of n line segments, namely the segments $(r_i e_0, r_i e_0 + e_i]$ for $i = 1, 2, \ldots, n$, and hence cannot contain a ray.

Editorial comment. Victor Klee showed in 1953 that in every infinite-dimensional real vector space X there is even a ubiquitous convex set containing no ray. A subset S of X is called ubiquitous if every point of X is an endpoint of some open segment contained in S. See pp. 107–108 of V. Klee, Convex sets in linear spaces III, *Duke Math. J.* 20(1953), 105–111. In the opposite direction S. K. Chung (Hong Kong) showed that a convex set in a finite-dimensional Hausdorff topological vector space contains a ray if its closure does.

Solved also by S. K. Chung (Hong Kong), Nir Cohen (Israel), A. A. Jagers (The Netherlands), Victor Klee, O. P. Lossers (The Netherlands), Charles Riley, John Henry Steelman, Richard B. Tucker, and the proposer.

Conjugacy Classes of Möbius Transformations

6589 [1989, 65]. *Proposed by Kevin Brown, Kent, WA.*

Given a positive integer g, let $n(g)$ denote the number of conjugacy classes of Möbius transformations each member of which generates a cyclic group of order g. Characterize those g greater than 6 for which $n(g) \equiv 1 \pmod 6$. (A Möbius transformation is a linear fractional transformation of the extended complex plane. Cf. Chapter II of Carathéodory's *Theory of Functions*, Volume I, Chelsea, 1954.)

Solution by the proposer. The numbers g that occur are all numbers of the form p^{2d} and $2p^{2d}$, where d is any positive integer and p is any prime congruent to 11 modulo 12 (the smallest such g is 121). For the proof we represent each Möbius

transformation by its corresponding matrix of determinant 1. Now it is well known that a Möbius transformation generates a cyclic group of order g if and only if Q, the square of its trace, is

$$Q = 4\cos^2\left(\frac{k\pi}{g}\right), \quad \text{with } (k, g) = 1.$$

(Cf. L. R. Ford, *Automorphic Functions*, McGraw-Hill, 1929, §8.) Also, two such transformations are similar if and only if they have the same squared trace. Hence the number of conjugacy classes of Möbius transformations that generate a cyclic group of order g is the number of distinct values of Q with $(k, g) = 1$. In terms of Euler's phi function this is $\varphi(g)/2$. Hence in terms of the decomposition of g into prime factors, we must find all

$$g = \prod p_i^{e_i} > 6$$

for which

$$\varphi(g)/2 = \tfrac{1}{2}\prod (p_i - 1)p_i^{e_i - 1} \equiv 1 \bmod 6.$$

This rules out more than one odd prime factor for g, and the problem is now reduced to a simple exercise (details omitted) in elementary number theory.

Solved also by Thomas McDonald and Allan Pedersen (Denmark).

The Evenly Spaced Integer Topology

6590 [1989, 65]. *Proposed by Jesús Ferrer, Oliva (Valencia), Spain.*

Let Z be the set of integers with the topology generated by the sets $a + kZ = \{a + kx: x \in Z\}$, where $a \in Z$, $k \in Z$, $k \neq 0$. (This topology, known as the Evenly Spaced Integer Topology, is studied in considerable detail in Steen and Seebach's *Counterexamples in Topology*.)

Suppose k and l are relatively prime positive integers with $k > 2$, and let $B_{k,l}$ be the set of all (positive) prime numbers congruent to l modulo k. Find the derived set of $B_{k,l}$, i.e., the set of accumulation points of $B_{k,l} = l + kZ$.

Solution by A. A. Jagers, Universiteit Twente, Enschede, The Netherlands. Let $B'_{k,l}$ denote the derived set of $B_{k,l}$. We prove that $B'_{k,l} = \{1\}$ if $l \equiv 1 \bmod k$ and $B'_{k,l} = \{-1\}$ if $l \equiv -1 \bmod k$, but that $B'_{k,l} = \varnothing$ otherwise. Let $1 + m\mathbb{Z}$ be a basis neighborhood of 1. Then $1 + km\mathbb{Z}$ is a smaller neighborhood of 1 containing infinitely many primes by Dirichlet's theorem. All these primes belong to $B_{k,l}$ if $l \equiv 1 \bmod k$. Hence 1 belongs to $B'_{k,l}$ if $l \equiv 1 \bmod k$. In the same way -1 belongs to $B'_{k,l}$ if $l \equiv -1 \bmod k$. On the other hand, if t is an integer other than -1, 0, or 1, then $t\mathbb{Z}$ is a neighborhood of t (and also of 0) that contains no (positive) primes except possibly t itself. Thus only 1 and -1 can occur as accumulation points of sets of primes. Finally, note that $l + k\mathbb{Z}$ is a closed neighborhood of $B_{k,l}$, so that any accumulation point of $B_{k,l}$ must belong to $l + k\mathbb{Z}$.

Editorial comment. Jagers also points out that $B'_{4,3} = \{-1\}$ (and, e.g., $B'_{5,3} = \varnothing$) instead of the incorrect $B'_{4,3} = \varnothing$ given in §58.5 of Steen and Seebach's *Counterexamples in Topology*.

Solved also by Michael Barr (Canada), Jesús Angel Hernández Isla (Spain), Nicholas Passell, and the proposer. One incorrect solution was received.

Integrals That Were Personal Friends of Ramanujan

6600 [1989, 366]. *Proposed by P. J. De Doelder, Technical University, Eindhoven, The Netherlands.*

If m is a positive integer, prove that

$$\int_0^{\pi/2} \frac{u^{2m}}{\sin u}\,du = 2(-1)^m(2m)!\lambda(2m+1) + 2(-1)^{m+1}(2m)!\sum_{j=0}^{m-1}\frac{(-1)^j(\pi/2)^{2j+1}}{(2j+1)!}\beta(2m-2j)$$

and

$$\int_0^{\pi/2} \frac{u^{2m-1}}{\sin u}\,du = 2(-1)^{m+1}(2m-1)!\sum_{j=0}^{m-1}\frac{(-1)^j(\pi/2)^{2j}}{(2j)!}\beta(2m-2j),$$

where

$$\lambda(s) = \sum_{k=0}^{\infty}\frac{1}{(2k+1)^s}, \qquad \beta(s) = \sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^s}.$$

Solution by Nathaniel Grossman, Department of Mathematics, UCLA. We note first that

$$\int_0^{\infty}\frac{u^p}{\sinh u}\,du = 2(p!)\lambda(p+1) \quad \text{and} \quad \int_0^{\infty}\frac{u^p}{\cosh u}\,du = 2(p!)\beta(p+1)$$

when p is a nonnegative integer. These two evaluations can be verified by expanding $1/\sinh u$ and $1/\cosh u$ into series of powers of e^{-u}, and then integrating term by term.

Let C denote the positively oriented, rectangular contour with corners at 0, $\pi/2$, $\pi/2 + iA$, and iA. By Cauchy's Theorem,

$$\int_C \frac{z^p}{\sin z}\,dz = 0.$$

Denote the sides of C by B, R, T, and L (for bottom, right, top, and left).

We let $A \to \infty$ and find by parametrizing the top segment of C that

$$\left|\int_T \frac{z^p}{\sin z}\,dz\right| = O\left(\frac{A^p}{\exp A}\right) = o(1).$$

Further obvious parametrizations yield

$$\begin{aligned}\int_R &= i\int_0^{\infty}\frac{(\pi/2+iu)^p}{\cosh u}\,du \\ &= \sum_{j=0}^{p}\binom{p}{j}\left(\frac{\pi}{2}\right)^j i^{p-j+1}\int_0^{\infty}\frac{u^{p-j}}{\cosh u}\,du \\ &= 2p!\,i^{p+1}\sum_{j=0}^{p}\frac{1}{j!}\left(-\frac{i\pi}{2}\right)^j\beta(p-j+1),\end{aligned}$$

and

$$\int_L = -i^p \int_0^\infty \frac{u^p}{\sinh u}\, du = -2p!i^p\lambda(p+1).$$

Hence,

$$\int_0^{\pi/2} \frac{u^p}{\sin u}\, du = -\int_L - \int_R$$

$$= 2p!i^p\lambda(p+1) - 2p!i^{p+1} \sum_{j=0}^{p} \frac{1}{j!}\left(-\frac{i\pi}{2}\right)^j \beta(p-j+1).$$

Now take $p = 2m$, an even, positive integer. We know that the integral is real-valued, so by equating real and imaginary parts we find that

$$\int_0^{\pi/2} \frac{u^{2m}}{\sin u}\, du = 2(-1)^m(2m)!\lambda(2m+1)$$

$$+ 2(-1)^{m+1}(2m)! \sum_{j=0}^{m-1} \frac{(-1)^j(\pi/2)^{2j+1}}{(2j+1)!}\beta(2m-2j),$$

as proposed, and also

$$\sum_{j=0}^{m} \frac{1}{(2j)!}\left(-\frac{\pi^2}{4}\right)^j \beta(2m-2j+1) = 0.$$

Note that in the latter identity every term is a rational multiple of π^{2m+1}.

When $p = 2m-1$, an odd positive integer, separation into real and imaginary parts gives

$$\int_0^{\pi/2} \frac{u^{2m-1}}{\sin u}\, du = 2(-1)^{m+1}(2m-1)! \sum_{j=0}^{m-1} \frac{(-1)^j(\pi/2)^{2j}}{(2j)!}\beta(2m-2j),$$

as proposed, and also

$$\lambda(2m) = \sum_{j=0}^{m-1} \frac{(-1)^j}{(2j+1)!}\left(\frac{\pi}{2}\right)^{2j+1} \beta(2m-2j-1).$$

Note that in the latter identity every term is a rational multiple of π^{2m}.

Editorial comment. A. A. Jagers gave a short elegant proof by bringing $f(u) = \log\tan(u/2)$ into the integrand via an integration by parts, expanding $f(u)$ into a Fourier series, and integrating termwise. C. Georghiou and Lynn R. Purser pointed out that somewhat more general results are given in Bruce C. Berndt, *Ramanujan's Notebooks* Part I, Springer-Verlag, New York, 1985, Chapter 9, Entry 14, p. 261.

Solved also by Jean Anglesio (France), C. Georghiou (Greece), A. A. Jagers (The Netherlands), Lynn R. Purser, Wolfgang Schuster (Germany), and the proposer.

POSTMORTEM COMMENTS ON ADVANCED PROBLEMS

Shapiro's Cyclic Inequality

4603 [1954, 571]. *Proposed by H. S. Shapiro, New York University.* (The present address of the proposer is: Royal Institute of Technology, Stockholm, Sweden.)

Given $x_i \geqq 0$, $i = 1, 2, \ldots, n$. Establish

$$\frac{x_1}{x_2 + x_3} + \frac{x_2}{x_3 + x_4} + \cdots + \frac{x_{n-1}}{x_n + x_1} + \frac{x_n}{x_1 + x_2} \geqq \frac{n}{2},$$

equality occurring only if all denominators are equal.

Remark by the editors. It is now known that the assertion of the problem is true only in the following two cases:

(i) n is even and $n \leq 12$

(ii) n is odd and $n \leq 23$.

For more detailed information see B. A. Troesch, The Validity of Shapiro's Cyclic Inequality, *Math. Comp.*, 53 (1989) 657–664 and the references given there.

Density of Pairs with Same Prime Factors

5735 [1970, 532; 1971, 680]. *Proposed by Paul Erdös and T. Motzkin, University College of Swansea, Wales.*

Denote by $F(n)$ the number of pairs $1 \leq a \leq b \leq n$ for which a and b have the same prime factors. Prove that $\lim_{n \to \infty} F(n)/n$ exists and is finite.

Remark by the editors. The solution published in 1971 was incorrect. It began with the erroneous assertion that if $n = r_1^{a_1} r_2^{a_2} \cdots r_k^{a_k}$, where the r_i are distinct prime numbers and the a_i are positive integers, then

$$F(n) - F(n-1) = a_1 a_2 \cdots a_k. \qquad (*)$$

The left-hand side of (*) is equal to the number of pairs (a, n) such that $1 \leq a \leq n$ and a and n have the same prime factors, while the right-hand side of (*) is equal to the number of such pairs (a, n) which satisfy the additional condition that $a|n$. For example, if $n = 50$, the left-hand side of (*) counts the four pairs (10, 50), (20, 50), (40, 50), and (50, 50), while the right-hand side of (*) counts only the two pairs (10, 50) and (50, 50). Thus the value given for $\lim F(n)/n$ in the published solution is too low.

For a correct proof of the assertion of the problem see H. N. Shapiro, Integer vectors with interprimed components, *Math. Comp.*, 27 (1973) 455–462, or Alan H. Stein, An application of Dirichlet series in several complex variables, *Comm. Pure Appl. Math.*, 28 (1975) 303–321.

REVIEWS

EDITED BY JOSEPH KONHAUSER
Department of Mathematics, Macalester College, St. Paul, MN 55105

The Emperor's New Mind. By Roger Penrose, Oxford University Press, New York, 1989, xiii + 466 pp.

MICHAEL BARR
Department of Mathematics and Statistics, McGill University, Montreal, Quebec, Canada H3A 2K6

This is a fascinating book, one that I recommend every mathematically sophisticated person read. I am a bit surprised by the number of people who have bought it, enough to put it on the bestseller list for several months. It certainly refutes the assertion in Hawking's *A Brief History of Time* that every formula contained in a book would reduce its sales by half.

The purpose of this book is to demonstrate that the human mind cannot be modeled by a Turing machine or other such mechanical device. The argument falls into two parts. The first part attempts to show that a human can best any machine by solving a problem or by proving a theorem that the machine cannot. If, in particular, the machine purported to be a model of that person's reasoning, there would be a contradiction. The more speculative second part describes how quantum theory provides a possible explanation for the first part.

This second part includes the most lucid exposition of quantum theory that I have ever read. The most lucid, but in the end not lucid at all. In fact, I am beginning to think that the phrase "lucid explanation of quantum theory" is simply an oxymoron. No matter: Penrose himself seems to agree. He certainly feels that there must be an explanation of reality of which quantum theory is simply the visible reflection.

He also gives a very clear (and this time convincing) description of time's arrow and the meaning of increasing entropy, as well as a description of how the universe began with minimum entropy.

The debate over whether human reasoning can be captured by a machine has a long history. The opening salvo was fired by J. R. Lucas [1961], although Lucas does not claim it was original with him, but that it was in the air at the time and he decided to formalize it and answer all the objections. He didn't answer quite all of them, a point to which I will return.

A second active participant in this debate has been J. R. Searle, who summarized his position in the January, 1990 *Scientific American* [Searle, 1990]. Searle posits the following axioms:

Axiom 1. Computer programs are formal (syntactic).

Axiom 2. Human minds have mental contents (semantics).

Axiom 3. Syntax by itself is neither constitutive of nor sufficient for semantics.

and, not surprisingly, concludes that human minds cannot be modeled by computer programs.

His main argument for Axiom 3 is a bizarre scenario known as the Chinese room experiment. It is hard to see what this experiment would show if it were actually carried out except that a person in a room full of instructions written on

3-by-5 cards could simulate a computer. Essentially, the purpose of the thought experiment is to allow Searle to assert that a computer cannot be said to understand a natural language.

This may be so; many people seem emotionally committed to this viewpoint. Many people in the last century (and a surprising number in this) are emotionally committed to the point of view that humans are not evolved from what they see as the lower animals. Theories of human exceptionality are rampant, but one should not assert them as axioms and expect to be taken seriously.

On the other hand, the outcome might surprise Searle, who doesn't appear to be aware of the dangers of supposing he knows what would happen in an experiment he hasn't done. Perhaps the man in the room, in the process of carrying out the experiment, would actually learn written Chinese purely from the syntax.

Let me offer a couple of arguments as to why there may, in fact, be no distinction between syntax and semantics. These arguments are not claimed to be conclusive; they only show that it is possible to hold the opposite point of view.

As any elementary school teacher will tell you, the only way to teach the meaning of a new word is to give several examples of its use *in context*. Dictionary definitions that don't show how the word is used are insufficient, and, once samples of use in enough contexts are given, turn out to be unnecessary. The extensive use of citations by the Oxford English Dictionary is another illustration of this fact.

But even more convincing evidence (to me at least) comes from my mathematical experience. In showing that every first order theory either has a model or can be demonstrated to be inconsistent, one uses the very sentences of the theory to build the model. In other words, there are always syntactic models and these are often the first ones one sees. Of course, one is happier with semantic models because one understands them better. But one understands them better because they are built from models of a theory that one understands better. The models of the earlier theory are either syntactic models of that theory or are derived, in their turn, from some more primitive theory. Follow it all the way down and it seems that all these semantic models may ultimately be based on syntax.

This doesn't prove that Searle is wrong, but his argument is seriously incomplete. And, as a matter of fact, Penrose does not consider that Searle's argument is conclusive, although he feels it has considerable force.

Lucas' argument is based on Gödel's theorem. It is identical in all relevant aspects to Penrose's, which I will discuss in considerable detail below. Oddly enough, Penrose fails to mention this fact, although he is aware of Lucas' article, referring to it in an end note to the last chapter, where he says, "See Lucas (1961) for an argument that Gödel's theorem implies non-computability and Good (1969), Benacerraf (1967) and Bowie (1982) for various counter-arguments." He also presents no discussion of the counter-arguments, which is even more surprising, since they refute his main argument as well.

In the library of the university that I visited, I found only Lucas' original article and one of the three refutations, that of Bowie [1982]. Since that is completely convincing, I will reproduce it in some detail.

Penrose bases his conclusions on two arguments, one of which is the Lucas argument. Before turning to that one, let us look at the one based on the undecidability of the "halting problem" for a Turing machine (or computer).

The argument begins on page 64 in a section titled, "How to outdo an algorithm" and the title is almost self-explanatory. To be perfectly definite, by a Turing machine T we will mean a machine (computer program, whatever) that takes as input a natural number n and either produces as output a natural number $T(n)$ or else runs forever, in which case we follow Penrose and write $T(n) = \square$.

It is well known that it is undecidable in general whether a Turing machine, given an input, will ever stop computing. For some machines, for some inputs, it is decidable, but there can be no general procedure that will always decide this question. No matter what decision procedure your algorithm uses, there is always some machine and input on which it will fail.

On the other hand, it is easy to imagine a machine H that takes as input Turing machines T (or you could use descriptions of Turing machines or even just Gödel numberings of such machines) and behaves as follows:

$$H(T, m) = \begin{cases} 1 & \text{if } T \text{ stops on input } m \\ 0 \text{ or } \square & \text{otherwise.} \end{cases}$$

This does not follow our definition of Turing machine because it takes as input a machine-natural number pair, but it can easily be coded to be a Turing machine. An example of such a machine is the one that, given input m, simulates the running of T on m and outputs 1 if T stops (and runs forever if T doesn't stop). Such a program H has the property that if $H(T, m) = 0$, then T will run forever on input m. On the other hand, it might happen that T runs forever on input m and so does $H(T, m)$. Since we cannot know if H is going to stop, we get no information on the pair (T, m) in that case.

Suppose now that we have such a program H and that $\{T_n\}$, $n \in \mathbf{N}$ is an enumeration of the Turing machines. The function f defined by

$$f(n) = 1 + T_n(n)H(T_n, n)$$

(and $f(n) = \square$ if the right-hand side is undefined) can be computed by some Turing machine so there is a k for which $f(n) = T_k(n)$.

Now Penrose argues that it is impossible for $H(T_k, k) = 1$, for that gives an outright contradiction. It is also impossible that $H(T_k, k) = 0$, for that implies that $T_k(k) = f(k) = \square$, while the right-hand side is 1. This is actually the case only if you have defined multiplication in such a way that $\square 0 = 0$. This is not the way most actual compilers arrange things (at the very least, it will depend on the order of evaluation), but it is perfectly possible to implement this kind of multiplication and I am willing to suppose that this is what Penrose had in mind. He now continues with the argument that although $H(T_k, k) = \square$ and so cannot decide if $T_k(k)$ terminates, *we* can see that it is impossible for $T_k(k)$ to terminate because then $H(T_k, k) = 1$, a contradiction. Thus we can decide a question that the algorithm H cannot and therefore we can beat any algorithm. (He isn't claiming that we can beat *every* algorithm simultaneously, but that claim isn't required for his argument.)

Where is the flaw? It is simply this: the whole chain of reasoning is based on the fact that H behaves as described above and *that we can prove it*. For it is knowing how H behaves that allows us to find a contradiction and to conclude that T_k runs forever. If H were an algorithm that *did* behave in this way, but for which that fact were unprovable, then the T_k constructed would still have the property that

$H(T_k, k)$ runs forever and so H cannot detect whether or not $T_k(k)$ is defined and *neither can we*. For our knowledge of the behavior of $H(T_k, k)$ is predicated on the partially effective nature of H and if we cannot know that, then we may not know the behavior of T_k either.

Let me now turn to the Penrose/Lucas argument from Gödel's theorem. In fact, the point at issue is really very close to the computational argument above. Gödel showed that for every set of axioms which is sufficiently strong for arithmetic there is what is now called a Gödel statement: a statement that is true in ordinary arithmetic but that cannot be proven from those axioms. Moreover, it is even possible to give an algorithm to construct such a sentence, given a set of axioms, much as T_k is constructed above to confound the machine H. But now, Penrose and Lucas argue, a person can readily see that this proposition is true since it was constructed to be true but unprovable in this system. Therefore there can be no axiomatization of the human thought processes.

It is a little harder to see what is wrong with this process. If we begin with a standard set of axioms, such as Peano arithmetic, then there is a Gödelian sentence and no one can doubt that the argument is correct and, assuming that those axioms are consistent (an important qualification as we will see), the sentence is true and is not provable from these axioms. But suppose we begin with a set of axioms for human thought (assuming such possible, which is what Penrose wants to refute). Or to be more precise, axioms for the thought processes of a single human being. It is not so clear that these axioms are consistent. In fact, it might be more reasonable to suppose the contrary. The Gödel sentence can be constructed as before, but now there is no reason to believe it true.

In a nutshell Bowie's argument is this: "Lucas must produce a Gödel sentence for *any* axiom set (and not only those for which he can prove consistency). If he believes all these Gödel sentences are true, then he is inconsistent." Later he adds, "Of course he would also be saved from inconsistency if, *without* a procedure, he produced Gödel sentences on just those occasions when he is challenged with a consistent machine, withholding them otherwise. But I, for one, do not believe he can do this." Nor do I believe that Penrose can.

You cannot escape this trap by sticking to axiom systems that can be proved consistent within themselves. For one thing, no such axiom system is strong enough to do arithmetic. But even if you lower your sights to relative consistency (relative to Peano arithmetic, say), it is still the case that you cannot prove consistency for any system as smart as yourself.

One way to summarize this argument is to point out that if I suppose that my reasoning is axiomatized by some set A of axioms, then I can still construct a Gödel sentence for A, but unless I can prove that A is consistent, I have no reason to believe that sentence is true. But if A is strong enough to do Peano arithmetic, then, according to another Gödel theorem, I must use a stronger axiom set than A to show that A is consistent. Thus I have no way of knowing if this Gödel sentence is true and so the contradiction that Lucas/Penrose hope to derive disappears. On page 116, Penrose dismisses the undecidability of consistency, which he describes as, "historically perhaps the most important part of his [Gödel's] argument", as irrelevant to the purposes of his book. To the contrary, this undecidability is the key to the unraveling of his argument.

On page 110, there is a curious statement, "The insight whereby we concluded that the Gödel proposition $P_k(k)$ is actually a true statement of arithmetic is an

example of a general type of procedure known to logicians as a *reflection principle*: thus by 'reflecting' upon the *meaning* of the axiom system and rules of procedure, and convincing oneself that these indeed provide valid ways of arriving at mathematical truths, one may be able to code this insight into further true mathematical statements that were not deducible from these very axioms and rules." This is nothing less than a chain that logicians use their reflections and insights, rather than formal reasoning to prove their theorems. This is an unbelievable howler of an error that is a libel on the world of logicians. There are subjective aspects in the choices of which sets of axioms you accept and the reflection principle (which is, in fact, a rather esoteric axiom of set theory that says that certain properties of very large infinite cardinal numbers have "reflections" in properties of more ordinary cardinals) is just as objective as any other axiom of logic.

One of the striking things about all these arguments, including Searle's, is that they are completely formal. It is possible to imagine a planet full of robots that we have programmed to study philosophy and logic, using some heuristic procedures that are even now being worked on, to have discovered all these arguments and to be drawing exactly the same conclusions. Except that if their programming were actually consistent, they would see the flaws in the arguments and would not thereby conclude that their own thought processes could not be modeled by a computer.

None of this refutes Penrose's main contention. It is entirely possible that human thought may not be modeled by a Turing machine. But nothing in the book establishes it either. As far as I am concerned, it is an empirical question that cannot be settled by any amount of armchair philosophizing, nor by any purely formal reasoning.

One thing not discussed is the possibility that a machine equipped with a random number generator might model a human mind. I don't regard this as a serious argument against Penrose's main thesis because such a hybrid is no longer a Turing machine.

There is an enormous amount of valuable material in this book. There is the quantum theory mentioned above and the discussion of Time's arrow. There is also a considerable amount of good mathematical exposition. For example, the discussion of both Turing machines and Gödel's theorem are probably the best elementary discussion of these matters I have seen. There is a good introduction to lambda calculus, to the Mandelbrot set and much, much more. The only real criticism is that there is no valid evidence brought to bear on its main thesis.

REFERENCES

1. G. L. Bowie, Lucas' number is finally up, *J. Philosophical Logic*, 11 (1982) 279–285.
2. J. R. Lucas, Minds, machines and Gödel, *Philosophy*, 36 (1961) 120–124. Reprinted in A. R. Anderson, Minds and Machines, 1964.
3. J. R. Searle, Is the brain's mind a computer program?, *Scientific American*, 262 (January, 1990) 26–31.

TELEGRAPHIC REVIEWS

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T: Textbook	P: Professional Reading	1–4: Semesters
C: Computer Software	L: Undergraduate Library	**: Special Emphasis
S: Supplementary Reading	13: Grade Level	??: Questionable

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

General, T, S, P, L**.** *Theorems and Counterexamples in Mathematics.* Bernard R. Gelbaum, John M.H. Olmsted. Problem Books in Math. Springer-Verlag, 1990, xxxiv + 305 pp, $42. [ISBN: 0-387-97342-7] Modelled after the authors' classic *Counterexamples in Analysis,* but larger in scope. This conversation for mathematicians amounts to a summing up of the canon: standard parts of algebra, analysis, geometry, logic, probability, set theory, and topology, including definitions, familiar notions and concepts, main results, examples and counterexamples, exercises, discussions. A guide (386 headings) and a glossary/index (50 pages) make the book easy to access. This book comes as close as any to being both an invaluable resource and a companionable book for bedtime reading. LCL

Education. *Projects to Enrich School Mathematics: Level 1.* Ed: Judith M. Trowell. NCTM, 1990, v + 168 pp, $14.50 (P). [ISBN: 0-87353-280-5] Independent study projects for mathematics students in grades 4-6. Companion volume to the two Projects books for middle and high school students. Eighteen topics, most not covered in the typical curriculum, challenge students to read, investigate, conjecture, and communicate. Good supplementary in-service text to give teachers ideas for implementing the NCTM *Curriculum and Evaluation Standards.* MW

Education, P, L. *Mathematical Knowledge of Japanese, Chinese, and American Elementary School Children.* James W. Stigler, Shin-Ying Lee, Harold W. Stevenson. NCTM, 1990, iv + 79 pp, $12 (P). [ISBN: 0-87353-294-5] Report of a three nation study of mathematical knowledge of first and fifth grade students in Sendai (China), Taipei (Taiwan), and Chicago (U.S.A.) on both computational and higher-order skills (word problems, estimation, visualization, mental calculation, etc.). The report confirms other studies by documenting deficiencies in American children that were "more serious than we expected." "There was no area in which [U.S. children] were competitive." Includes full texts of the tests with item responses for each nation in the study. LAS

Education, P, L. *Precollege Science and Mathematics Teachers: Monitoring Supply, Demand, and Quality.* Eds: Dorothy M. Gilford, Ellen Tenenbaum. National Academy Pr, 1990, x + 256 pp, (P). [ISBN: 0-309-04197-X] A carefully reasoned analysis of the incredible shortcomings of data concerning supply, demand, and quality of science and mathematics teachers: from definition of terms (how can there be a shortage since every class has a teacher?) to collection of data (how can one measure the reserve pool of experienced teachers who are not currently teaching?), only one type of recommendation could emerge: collect better data and conduct further research. No surprises here. LAS

Education, P, L. *Proceedings of the Conference on Technology in Collegiate Mathematics: Teaching and Learning with Technology.* Eds: Franklin Demana, Bert K. Waits, John Harvey. Addison-Wesley, 1991, 352 pp, (P). [ISBN: 0-201-54643-4] Nearly 100 papers on ways college faculty are using calculators and computers to teach mathematics. Courses range from high school algebra to abstract algebra, technology from graphing calculators to *Mathematica.* Approximately half of the volume is devoted to calculus. An inviting potpourri of ideas and experiences at the forefront of curricular innovation. LAS

Education, P. *Proceedings of the National Conference on Women in Mathematics and the Sciences.* Eds: Sandra Z. Keith, Philip Keith (Dept. of Math., St. Cloud St. U., St. Cloud, MN 56301), 1990, vi + 249 pp, (P). 75 papers and 20 abstracts from the second national conference on women in mathematics and science. Categories into which papers are grouped reflect the diverse issues covered in the volume: public affairs, outreach, model programs, "the glass ceiling," research, courses and curriculum, teacher support, remedial programs, networking. An immense and invaluable resource and data base for national networking on an issue of increasing national importance. LAS

Education, P. *Math/Science Education Action Conference Report.* US Department of Energy, 1990, 81 pp, (P). Yet another report on goals in math-

ematics and science—this one from the U.S. Department of Energy (not, as one might imagine from the Department of Education). Stresses five predictable goals for students, teachers, underrepresented groups, public literacy, and competitive work force. Commits DoE to an active role (via its personnel and its laboratories) in efforts to achieve these goals. Participants include many of the nation's leading scientists, educators, and policy makers. LAS

History, P, L. *The Legacy of John von Neumann.* Eds: James Glimm, John Impagliazzo, Isadore Singer. Proc. of Sympos. in Pure Math., V. 50. AMS, 1990, vii + 334 pp, $56. [ISBN: 0-8218-1487-7] Proceedings of a June 1988 conference at Hofstra University—one of the special AMS centennial events—celebrating the mathematical and scientific contributions of John von Neumann. Seventeen overview papers describing current work in some of the many fields stimulated by von Neumann (ergodic theory, operator theory, mathematical physics, computing, turbulence, neural networks, game theory, automata theory, etc.) are preceded by several personal reminiscences, including special tributes by von Neumann's daughter Marina Whitman and his brother Nicholas Vonneuman. LAS

Logic, T(15-17: 1), L. *Mathematical Logic for Computer Science.* Lu Zhongwan. World Scientific, 1989, ix + 248 pp, $47. [ISBN: 9971-502518] Concise text covers propositional logic, first-order logic, constructive logic, modal logic. Uses natural deduction approach. Prerequisites are basic set theory and mathematical maturity. Does not discuss applications of logic to computer science. KES

Logic, P. *Logic and Computation.* Ed: Wilfried Sieg. Contemporary Math., V. 106. AMS, 1990, xiv + 297 pp, $37 (P). [ISBN: 0-8218-5110-1] Proceedings of a workshop held at Carnegie Mellon University, June 30-July 2, 1987. An important aspect is the emphasis on feasibility in practice, rather than simply computability in principle. LCL

Combinatorics, P. *Coding Theory and Design Theory, Part I: Coding Theory.* Dijen Ray-Chaudhuri. Inst. for Math. & its Applic., V. 20. Springer-Verlag, 1990, 239 pp, $29. [ISBN: 0-387-97228-5] Part of the proceedings of a workshop held at the Institute of Mathematics and Its Applications, University of Minnesota, in June 1988. The papers represent research at the cutting edge of current work in this area. LCL

Discrete Mathematics, T(14: 1, 2). *Foundations of Discrete Mathematics, Second Edition.* Albert D. Polimeni, H. Joseph Straight. Brooks/Cole, 1990, xv + 519 pp, $44.75. [ISBN: 0-534-12402-X] Appropriate for a sophomore-level course in which students are introduced to abstract mathematics. Contains topics which are basic to the study of computer science, however it can be used in a course which is exclusively devoted to mathematics. Includes logic, set theory, number theory, mathematical induction, relations, functions, combinatorics, graph theory, algebraic structures, and matrices. Changes from the previous edition included more combinatorics, greater emphasis on algorithms and directed graphs, and more routine exercises. RH

Number Theory, T(18: 2), P. L. *Lattice Points.* Ekkehard Krätzel. Math. & Its Applic. Kluwer Academic, 1988, 320 pp, $99.50. [ISBN: 90-277-2733-3] The standard problem in the theory of lattice points is to estimate the number of such points contained in large closed domains of Euclidian space, a problem closely related to that of estimating exponential sums $\Sigma_{n\in D} \exp[2\pi i f(n)]$ where f is a real function and n is a lattice point of a domain D in p-dimensional Euclidean space. This book is an introduction to methods used to arrive at these estimations. Included are applications to the circle and sphere problem, Dirichlet's divisor problem, and the distribution of powerful numbers (a number is *powerful of type k* if all of its prime factors divide it at least to the k-th power). MPR

Number Theory, T*(16-17), S, L.** *Primes of the Form $x^2 + ny^2$: Fermat, Class Field Theory, and Complex Multiplication.* David A. Cox. Wiley, 1989, xi + 351 pp, $42.95. [ISBN: 0-471-50654-0] A unique and sensational book. Its goal is to answer the question: Given a positive integer n, which primes have the form $x^2 + ny^2$ where x and y are integers? The complete answer leads the reader on a wonderful excursion through the core of classical and modern algebraic number theory—quadratic forms, reciprocity laws, class field theory, elliptic and modular functions. There are many rich exercises, and the text is loaded with interesting, historical remarks. SG

Number Theory, T(17-18: 1), S, P, L. *Introductory Lectures on Siegel Modular Forms.* Helmut Klingen. Stud. in Adv. Math., V. 20. Cambridge University Pr, 1990, x + 162 pp, $39.50. [ISBN: 0-521-35052-2] An introduction to Siegel modular forms for graduate students or very advanced undergraduates with one semester of complex analysis and basic knowledge of algebra, number theory, and topology. Includes some historical discussions and extensive references. Chapters cover the modular group, basic facts on modular forms, large weights, small weights, modular functions and Dirichlet series. Written at an advanced, but readable, level. No exercises. SB

Algebra, P. *Algebraic Analysis, Geometry, and Number Theory: Proceedings of the JAMI Inaugural Conference.* Ed: Jun-Ichi Igusa. Johns Hopkins University Pr, 1989, xi + 417 pp, $45. [ISBN: 0-8018-3841-X] Proceedings of the Inaugural Conference for the Japan-US Mathematics Institute held at Johns Hopkins University from May 16-19, 1988. Includes seventeen articles. RH

Calculus, T(13: 1). *Calculus with Applications to the Management, Life, and Social Sciences.* Lawrence E. Spence, Charles Vanden Eynden. Scott Foresman, 1990, xvi + 701 pp. [ISBN: 0-673-18838-8] Introduction to integral and differential calculus assuming only high school algebra. Includes single

chapters on probability, sequences and series, and multivariate calculus. Applications include single-server queues, the method of least squares, and simple and compound interest. RWJ

Real Analysis, T(18: 1, 2), P. *Lanzhou Lectures on Henstock Integration.* Lee Peng-Yee. Ser. in Real Analysis, V. 2. World Scientific, 1989, viii + 179 pp, $48; $22 (P). [ISBN: 9971-50-891-5; 9971-50-892-3] The Henstock integral is more general, and in some respects simpler, than the integrals of Riemann, Lebesgue, Denjoy, etc. This monograph, based on a 1985 series of lectures at Lanzhou, China begins with basic aspects of Henstock integral theory and proceeds to convergence theory, connections with other integrals, and generalizations. Style is comparatively leisurely, textbook-like, self-contained; there are many examples, but no exercises. PZ

Real Analysis, T(16-17: 1, 2), S. *Principles of Real Analysis, Second Edition.* Charalambos D. Aliprantis, Owen Burkinshaw. Academic Pr, 1990, xii + 295 pp, $39.95. [ISBN: 0-12-050255-0] *Second Edition* of this well-written introduction to measure and integration. The previous edition (TR, November 1982) contained ample exercises and examples. This edition adds review problems at the end of each chapter (this addition is the main difference between printings). KS

Real Analysis, S*(15), L*. *Theory and Application of Infinite Series.* Konrad Knopp. Transl: R.C.H. Young. Dover, 1990, xii + 563 pp, $12.95 (P). [ISBN: 0-486-66165-2] Typical Dover reprint of a classic. In addition to material on series generally covered in a first analysis course are chapters on products, series of complex numbers, Fourier series, divergent series, and asymptotic expansions. Probably too concise for a textbook (and there are only limited exercises), but just right for a readable reference. GG

Complex Analysis, T(18: 1, 2), P. *Fix-Points and Factorization of Meromorphic Functions.* Chi-Tai Chuang, Chung-Chun Yang. World Scientific, 1990, viii + 227 pp, $46. [ISBN: 981-02-0008-0] English translation of a 1988 monograph, in Chinese, on theory and applications of meromorphic functions. In four long chapters and an appendix: the first two chapters review Nevanlinna theory, normal families, and basics of fix-point theory; the last two chapters introduce factorization theory and relations with fix-point theory. A variety of open research problems are surveyed. PZ

Complex Analysis, T(18), S, P. *An Introduction to Complex Analysis in Several Variables.* Lars Hörmander. Math. Library, V. 7. North-Holland (US Distr: Elsevier Science), 1990, xii + 254 pp, $71.75. [ISBN: 0-444-88446-7] *Third Edition* of this well-known, self-contained exposition of complex analysis from point of view of the theory of partial differential equations. The additions stress results where the analytic techniques based on the solution of the $\overline{\partial}$ Neumann problem are of particular importance (*First Edition,* TR, January 1967; *Second Edition,* TR, August-September 1974). KS

Partial Differential Equations, P. *Symmetries of Partial Differential Equations: Conservation Laws—Applications—Algorithms.* Ed: A.M. Vinogradov. Kluwer Academic, 1989, vi + 242 pp, $129. [ISBN: 0-7923-0594-9] Reprinted from *Acta Applicandae Mathematicae,* Vol. 15, Nos. 1 and 2, and Vol. 16, Nos. 1 and 2. Divided into three parts. Part I considers methods of finding all higher symmetries and conservation laws arising from a given differential equation. Part II considers relationships between symmetry and integrability of systems of differential equations. Part III addresses computer related issues. SM

Partial Differential Equations, P, L. *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations.* Eds: Tony F. Chan, *et al.* SIAM, 1990, xx + 491 pp, $58.50 (P). [ISBN: 0-89871-253-X] Proceedings of a symposium held in Houston in March 1990. Four parts: theory, algorithms, parallel implementation, and applications. SM

Numerical Analysis, P. *Computational Solution of Nonlinear Systems of Equations.* Eds: Eugene L. Allgower, Kurt Georg. Lect. in Appl. Math., V. 26. AMS, 1990, xix + 762 pp, $215. [ISBN: 0-8218-1131-2] Proceedings of an AMS-SIAM Summer Seminar held July 18-29, 1988 at Colorado State University in Fort Collins. The coverage is quite broad, including topics from complementarity theory and differential equations. SM

Functional Analysis, T(17: 1). *Essential Results of Functional Analysis.* Robert J. Zimmer. Lect. in Math. Ser. University of Chicago Pr, 1990, ix + 157 pp, $14.95 (P); $34.95. [ISBN: 0-226-98338-2; 0-226-98337-4] A concise and easy-to-read presentation of fundamental results in functional analysis. Variety of topics including convexity, fixed point theorems, compact operators, compact groups, spectral theory, ergodic theory, C^*-algebras, Fourier transforms, and many more. Exercise sets follow each chapter. MLR

Analysis, T(15-17), S, P, L. *Special Functions.* Z.X. Wang, D.R. Guo. Transl: D.R. Guo, X.J. Xia. World Scientific, 1989, xvi + 695 pp, $75; $38 (P). [ISBN: 9971-50-659-9; 9971-50-667-X] A compendium of the various principal special functions in common use and their basic properties and derivations. Some of the less important formulas are put into the exercises, together with relevant hints for the more difficult ones. KS

Algebraic Geometry, P. *Néron Models.* Siegfried Bosch, Werner Lütkebohmert, Michel Raynaud. Ser. of Modern Surveys in Math., Band 21. Springer-Verlag, 1990, x + 325 pp, $85.50. [ISBN: 0-387-50587-3] The definitive treatment of Néron models in the setting of schemes. The authors provide motivation for the theory and some background material from algebraic geometry. Other topics include birational group laws, descent, the Picard functor, and Jacobians. SG

Algebraic Geometry, P. *Lecture Notes in Mathe-*

matics-1420: Real Analytic and Algebraic Geometry. Eds: M. Galbiati, A. Tognoli. Springer-Verlag, 1990, 366 pp, $35.90 (P). [ISBN: 0-387-52313-8] A collection of twenty-one research papers presented at a 1988 conference in Trento. SG

Differential Geometry, P. *Geometric and Topological Invariants of Elliptic Operators.* Ed: Jerome Kaminker. Contemporary Math., V. 105. AMS, 1990, xi + 297 pp, $37 (P). [ISBN: 0-8218-5112-8] Proceedings of a conference held July 1988 at Bowdoin College. Contains twelve papers. MLR

Differential Geometry, T?, S*(16-17). *Solid Shape.* Jan J. Koenderink. MIT Pr, 1990, xiv + 699 pp, $65. [ISBN: 0-262-11139-X] A largely expository text on the differential geometry of submanifolds of three-dimensional space, written from the perspective of an engineer "hooked on differential geometry." Probably not ideal for a main text (no exercises, few computations, often vague definitions and theorem statements), but contains some marvelous intuition that makes it well worth reading. JO

Differential Geometry, T(17-18: 1, 2). *Modern Differential Geometry for Physicists.* Chris J. Isham. Lect. Notes in Physics, V. 32. World Scientific, 1989, vii + 182 pp, $42; $22 (P). [ISBN: 9971-50-956-3; 9971-50-957-1] Lecture notes from a first-year course for physics graduate students. Chapters on manifolds, Lie groups, and fibre bundles. Author describes exposition as "pithy," and in need of supplementary discussion. He should be heeded. JO

Geometry, T?(16-18), S?, P. *Fractal Geometry: Mathematical Foundations and Applications.* Kenneth Falconer. Wiley, 1990, xxii + 288 pp, $33.95. [ISBN: 0-471-92287-0] Begins with various notions of dimension and methods for their calculation. Foundational part concludes by examining geometrical properties of fractals. The second part, on examples and applications, includes self-similar and self-affine sets, examples from number theory and pure mathematics, dynamical systems, Julia sets, random fractals, and Brownian motion. Aimed at researchers, this book contains exercises and thus may also be useful as a course text. KS

Algebraic Topology, P. *Braid Group, Knot Theory and Statistical Mechanics.* Eds: C.N. Yang, M.L. Ge. Adv. Ser. in Math. Physics, V. 9. World Scientific, 1989, vii + 329 pp, $68; $38 (P). [ISBN: 9971-50-828-1; 9971-50-833-8] Nine articles on these three topics illustrating the fascinating interplay among them. In particular, one point of focus is the role of the Yang-Baxter equations as the key to solvability in each of these areas. A warning: the typesetting in several articles is somewhat primitive. MPR

Operations Research, P, L. *Twenty-five Years of Operations Research in the Netherlands: Papers Dedicated to Gijs de Leve.* Eds: Jan Karel Lenstra, Henk Tijms, Ton Volgenant. CWI Tracts, V. 70. Mathematisch Centrum, 1989, 181 pp, Dfl. 48 (P). [ISBN: 90-6196-385-0] A collection of sixteen papers contributed by current and former students and colleagues of Gijs de Leve on the occasion of his silver jubilee as Professor at the University of Amsterdam. A wide variety of topics are represented, with most papers dealing with some aspect of either a Markov process or optimization on a network. SM

Dynamical Systems, P. *Bifurcation and Chaos in Simple Dynamical Systems.* J. Awrejcewicz. World Scientific, 1989, viii + 126 pp, $26. [ISBN: 9810-200382] Collection of the author's results, including an analytical method of solving the Hopf bifurcation problem, a numerical method for studying bifurcation behavior of nonlinear differential equations, and an analytical condition for special types of Hopf bifurcations. JO

Dynamical Systems, T*(16: 2), L. *Chaotic Dynamics of Nonlinear Systems.* S. Neil Rasband. Wiley, 1990, x + 230 pp, $39.95. [ISBN: 0-471-63418-2] A very nice text intended primarily for physics students interested in the mathematics of chaos. Most of the "classical" topics are represented, e.g., one-dimensional maps, universality, fractal dimension, differential dynamics, conservative dynamics, and measures of chaos. The emphasis throughout is on numerical calculations (on PC's) to illustrate the key ideas and theorems. Plenty of exercises and examples. Also includes an extensive bibliography and a glossary (very useful!). MPR

Dynamical Systems, S(16-18), P, L. *Elementary Symbolic Dynamics and Chaos in Dissipative Systems.* Hao Bai-lin. World Scientific, 1989, xv + 460 pp, $78; $38 (P). [ISBN: 9971-50-682-3; 9971-50-698-X] A physics-oriented introduction to dynamical systems, emphasizing symbolic dynamics. Lots of detailed examples. BC

Control Theory, P. *Optimal Radar Tracking Systems.* George Biernson. Wiley, 1990, xiv + 594 pp, $64.95. [ISBN: 0-471-50673-7] A detailed engineering study of the Altair Radar System located on the Kwajalein Atoll. The emphasis is on a practical application of optimal control theory, particularly Kalman filtering. The presentation should be of interest to any engineer or mathematician concerned with any type of target tracking. SM

Control Theory, T(18: 1, 2), P, L. *Nonlinear Dynamical Control Systems.* Henk Nijmeijer, Arjan van der Schaft. Springer-Verlag, 1990, xiii + 467 pp, $49.95. [ISBN: 0-387-97234-X] A graduate-level textbook applying geometric control theory to nonlinear systems. Many examples of both discrete and continuous time systems. Also included is a self-contained discussion of the necessary differential geometry. The book is well-written with plenty of exercises. SM

Control Theory, P. *Lecture Notes in Control and Information Sciences-141: Root Clustering in Parameter Space.* S. Gutman. Springer-Verlag, 1990, viii + 153 pp, $27.10 (P). [ISBN: 0-387-52361-8] Certain control problems require the eigenvalues of a matrix to "cluster" in a given part of the complex plane. A theoretic treatment, with lots of detailed examples. BC

Elementary Statistics, T*(13-14: 1, 2). *Ele-*

mentary Statistics in a World of Applications, Third Edition. Ramakant Khazanie. Scott Foresman, 1990, 716 pp, $32.95. [ISBN: 0-673-16751-8] Revision of the author's 1986 *Second Edition (First Edition,* TR, April 1979; *Second Edition,* TR, November 1986). Most significant change is the inclusion of Minitab labs at the end of each chapter with detailed explanations of the commands. New exercises based on real data and some new material (e.g., box plots, Bayes' theorem, sampling techniques) have also been added, and review exercises have replaced the chapter tests. RSK

Elementary Statistics, T? *Statistics in Scientific Investigation: Its Basis, Application, and Interpretation.* Glen McPherson. Texts in Statistics. Springer-Verlag, 1990, xxvi + 666 pp, $49.95. [ISBN: 0-387-97137-8] Applied statistics textbook assuming as pre-requisites high school algebra and understanding of scientific modelling. The book includes much discussion about statistics but models few illuminating examples. Ill-suited for those with just an algebra background and no scientific experience. TH

Elementary Statistics, T(13-16: 1), L. *Understanding Statistics in the Social Sciences.* William Ray Arney. WH Freeman, 1990, xvii + 524 pp, $19.95 (P). [ISBN: 0-7167-2006-X] Delightful introduction to the concepts and language of statistics with a bit of history interwoven. Covers basic probability, descriptive statistics, hypothesis testing (including tests for categorical data, and one- and two-way ANOVA), point and interval estimation, measures of association, and linear regression models. Good exercises and illustrative examples. RWJ

Computational Statistics, P. *MAXDATA: A Time Series Database System.* Wilhelm A. Hennerkes. Springer-Verlag, 1990, xiii + 150 pp, $35 (P). [ISBN: 0-387-52209-3] Handbook describing the MAXDATA database system, designed to be a user-friendly software package for personal computers for the analysis and management of large economic data sets. RSK

Statistics, T(15-17: 1), S, L. *Finding Groups in Data: An Introduction to Cluster Analysis.* Leonard Kaufman, Peter J. Rousseeuw. Ser. in Prob. & Math. Stat. Wiley, 1990, xiv + 342 pp, $49.95. [ISBN: 0-471-87876-6] Describes some robust partitional and hierarchical (divisive and agglomerative) clustering methods of general applicability. Prefaced by guidelines for the choice of a method and instructions on how to prepare data for clustering. Includes detailed instruction on how to use the author's Fortran software, computer output of several examples, and about 80 exercises. RWJ

Statistics, P. *Lecture Notes in Statistics-62: Testing Problems with Linear or Angular Inequality Constraints.* Johan C. Akkerboom. Springer-Verlag, 1990, xii + 291 pp, $35.90 (P). [ISBN: 0-387-97232-3] This monograph is the published doctoral thesis of the author. It presents a likelihood ratio method for testing problems with an alternative hypothesis restricted by linear inequality constraints. The monograph focuses on the practical application of Pincus's idea to replace the linear inequalities with one angular inequality, a circular cone, and then to apply a likelihood ratio test. JJ

Statistics, T(18: 1) *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses.* Wayne Nelson. Ser. in Prob. & Math. Stat. Wiley, 1990, xiv + 601 pp, $69.95. [ISBN: 0-471-52277-5] Excellent applied practical text on accelerated life (reliability) testing, the process of testing products at higher than normal stress or usage levels in order to reduce test time, and then relating the accelerated failure distributions back to normal time/usage levels. Text contains many real data sets and specific examples analyzing these data. The newest developments in the accelerated testing field are covered. Analysis methods from simple graphical techniques to advanced computer programs are presented. Life-stress models include inverse power, Eyring's, Arrhenius', Zhurkov's. Complete data analyzed with least squares, censored data with maximum likelihood methods. Test plans are discussed, as are models for handling data with varying stress. Product data analyzed from electronics, insulations, semiconductors, plastics, metals. JJ

Statistics, S(18), P. *Spline Models for Observational Data.* Grace Wahba. CBMS-NSF Reg. Conf. Ser. in Appl. Math., V. 59. SIAM, 1990, xii + 169 pp, $24.75 (P). [ISBN: 089871-244-0] Monograph from a 1987 lecture series held in Columbus, Ohio. Concentrates on work by the author and associates. Develops univariate smoothing splines, thin plate splines, splines on the sphere, etc., using reproducing kernel Hilbert space theory. Convergence properties, confidence intervals, parameter selection. References to numerical methods and software. TH

Statistics, P. *Lecture Notes in Statistics-63: Estimation in Semiparametric Models: Some Recent Developments.* Johann Pfanzagl. Springer-Verlag, 1990, 112 pp, $17.60 (P). [ISBN: 0-387-97238-2] Constructs asymptotically optimal sequences of estimates for functionals of distributions having a finite dimensional parameter and which admit a sufficient statistic for the nuisance parameter. RWJ

Statistics, T*(16-17: 1), P, L. *Categorical Data Analysis.* Alan Agresti. Ser. in Prob. & Math. Stat. Wiley, 1990, xv + 558 pp, $49.95. [ISBN: 0-471-85301-1] Divided into four main parts: descriptive and inferential methods for two-way contingency tables; basics of model building, emphasizing loglinear and logit models, for two-way and multi-way tables; applications and generalizations of these models; and the theoretical foundations for modeling categorical data. Very good problem sets, both theoretical and applied. Extensive set of references. RSK

Statistics, P. *Distributions of Test Statistics: Exact & Asymptotic, Null & Non-Null, Methods/Comparisons/Research Frontiers.* Ed: A.M. Mathai. Amer. Ser. in Math. & Management Sci., V. 18. American Sciences Pr, 1989, 182 pp, $89.75 (P). [ISBN: 0-935950-20-6] Collection of eight re-

view articles concerned with tests of hypotheses on one or more multivariate normal populations. Centers on distributions connected with likelihood ratio test statistics. (Also issued as a special issue of the *American Journal of Mathematical and Management Sciences,* Volume 9, Nos. 1 and 2.) Note price! RSK

Artificial Intelligence, T(17-18: 2), S, P, L. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms.* Richard E. Neapolitan. Wiley, 1990, xiii + 433 pp, $44.95. [ISBN: 0-471-61840-3] Graduate-level text on the use of probability and graph theory (networks) in expert systems. Includes exercises and examples. BC

Artificial Intelligence, S, P. *Lecture Notes in Artificial Intelligence-419: A Methodology for Uncertainty in Knowledge-Based Systems.* Kurt Weichselberger, Sigrid Pöhlmann. Springer-Verlag, 1990, 132 pp, $16.50 (P). [ISBN: 0-387-52336-7] In the area of expert systems, it is important for the system designer to be able to represent and manipulate incomplete, missing, or uncertain information. This text reviews the current research in the area of artificial intelligence, and presents one specific methodology for dealing with this problem, based on probability theory. The authors then give an example of how their theory can be applied to expert system in the area of diagnostic medicine. GMS

Computer Science, T(16-17: 1). *Data Analysis for Data Base Design, Second Edition.* D.R. Howe. Edward Arnold (US Distr: Routledge, Chapman & Hall), 1989, xiii + 317 pp, $27.50 (P). [ISBN: 0-7131-3688-X] This edition is designed to make data analysis techniques more readily available to students of systems analysis. The most significant change from the previous edition is the addition of a chapter on the SQL relational data base language. The book is divided into four parts: data base management systems, relational modelling, entity-relational modelling, implementation. Case studies are used, and problems and assignments given. JJ

Applications, P, L. *Environmental Effects on Bus Durability.* Sandra L. Arlinghaus, John D. Nystuen. Monograph, No. 11. Institute of Mathematical Geography, 1989, vi + 104 pp, $15.95 (P). This monograph considers the effects of climate, terrain, congestion, and the combination of all three factors on the durability of buses serving public transit agencies in the United States. For each factor the work discusses, in detail, how the almost 200 cities studied are classified into "peer" groups, and what implications these classifications have for maintenance of the bus fleet. An interesting application of statistical and geographic data analysis. SM

Applications (Behavioral Science), P, L*. *The Anthropology of Numbers.* Thomas Crump. Stud. in Social Anthropology. Cambridge University Pr, 1990, x + 200 pp, $49.50. [ISBN: 0-521-38045-6] The first scholarly work to examine the role of numbers (including numeracy and numerology) from a broad cultural perspective. Detailed logical, psychological, and linguistic foundations provide a basis for an exploration of the role numbers play in various cultures' cosmology, economy, music, games, politics, architecture, and language. Profusely illustrated with examples from societies past and present—especially from the Far East where the author lived for several years. Extensive end notes; thorough bibliography; comprehensive index. A marvelous resource for anyone interested in the role of mathematics in culture. LAS

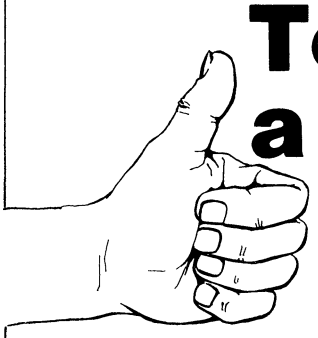
Applications (Communication Theory), S(17-18), P. *Time Series Analysis, Identification and Adaptive Filtering, Second Edition.* Daniel Graupe. Robert E Krieger, 1989, xvii + 420 pp, $49.50. [ISBN: 0-89464-315-0] Self-contained text on applications of time series analysis to signal identification and adaptive filtering of unknown signals from unknown noise. This edition adds a chapter on time series with time-varying parameters. (*First Edition,* TR, February 1985.) BC

Applications (Physics), L. *Fractal Growth Phenomena.* Tamás Vicsek. World Scientific, 1989, xi + 355 pp, $67. [ISBN: 9971-50-442-X] Basic concepts in the study of fractal growth (concentrating on geometric aspects). Part I serves as an introduction explaining the basics of fractal geometry: fractal dimension, measure, and many examples (experimental and computer generated). Parts II and III apply fractal geometry to the physics of growth phenomena. MLR

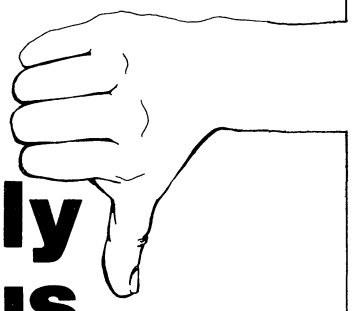
Applications (Social Science), T, S(15-17), L. *Understanding Quantitative History.* New Liberal Arts Ser. Loren Haskins, Kirk Jeffrey. MIT Pr, 1990, xxv + 366 pp, $27.50. [ISBN: 0-262-08190-3] Case studies from published historical research (primarily American history) organized in chapters to introduce prospective historians and students of other social sciences to basic tools of descriptive statistics: tables, charts, graphs, means and medians, samples, standard deviation, scatter charts, and regression. Requires only first year high school algebra; uses no formulas. Numerous exercise questions invite the reader to analyze the case studies with care. (Answers to all questions are provided at the end of each chapter.) One of the Sloan "New Liberal Arts" textbooks intended to infuse greater attention to technology and applied mathematics into traditional liberal arts disciplines. LAS

Reviewers

SB: Steve Benson, St. Olaf; BC: Barry Cipra, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; RWJ: Roger W. Johnson, Carleton; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; JO: Jeff Ondich, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; KS: Karen Saxe, St. Olaf; GMS: G. Michael Schneider, Macalester; KES: Kay E. Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; PZ: Paul Zorn, St. Olaf.



Toward a Lean and Lively Calculus

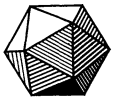


Report of the Conference/Workshop to Develop Curriculum and Teaching Methods for Calculus at the College Level, Ronald G. Douglas, Editor.
MAA Notes #6

249 pp., 1987, Paperbound, ISBN-0-88385-056-7
Catalog Number - NTE-07
Price: $12.50

Should calculus be taught differently? Can it? Common wisdom says "no"—which topics are taught, and when, are dictated by the logic of the subject and by client department. The surprising answer from a four-day Sloan Foundation-sponsored conference on calculus instruction, chaired by Ronald Douglas, is that significant change is possible, desirable, and necessary. Meeting at Tulane University in New Orleans in January, 1986, a diverse and sometimes contentious group of twenty-five faculty, university and foundation administrators, and scientists from client departments, put aside their differences to call for a leaner, livelier, more contemporary course, more sharply focused on calculus's central ideas and on its role as the language of science.

This volume contains the results of that conference and the papers presented to the conferees. These are certain to be the point of departure and basis for efforts to strengthen and reshape calculus in the next decade.



Order from
The Mathematical Association of America
1529 Eighteenth St., NW
Washington, DC 20036

Exponential Sums and Differential Equations

Nicholas M. Katz

This book is concerned with the theory of linear differential equations in one complex variable with polynomial coefficients, and the theory of one parameter families of exponential sums over finite fields. After reviewing results from representation theory, the book discusses results about differential equations and their differential galois groups (G_{gal}) and one-parameter families of exponential sums and their geometric monodromy groups (G_{mono}). The final part of the book is devoted to comparison theorems relating G_{gal} and G_{mono} of suitably "corresponding" situations, which provide a systematic explanation of the remarkable "coincidences" found "by hand" in the hypergeometric case.

Annals of Mathematics Studies, 124

Paper: $22.50 ISBN 0-691-08599-4
Cloth: $65.00 ISBN 0-691-08598-6
In Japan order from United Publishers Services

Automorphic Representations of Unitary Groups in Three Variables

Jonathan D. Rogawski

The purpose of this book is to develop the stable trace formula for unitary groups in three variables. The stable trace formula is then applied to obtain a classification of automorphic representations. This work represents the first case in which the stable trace formula has been worked out beyond the case of SL (2) and related groups. Many phenomena which will appear in the general case present themselves already for these unitary groups.

Annals of Mathematics Studies, 123

Paper: $19.95 ISBN 0-691-08587-0
Cloth: $60.00 ISBN 0-691-08586-2
In Japan order from United Publishers Services



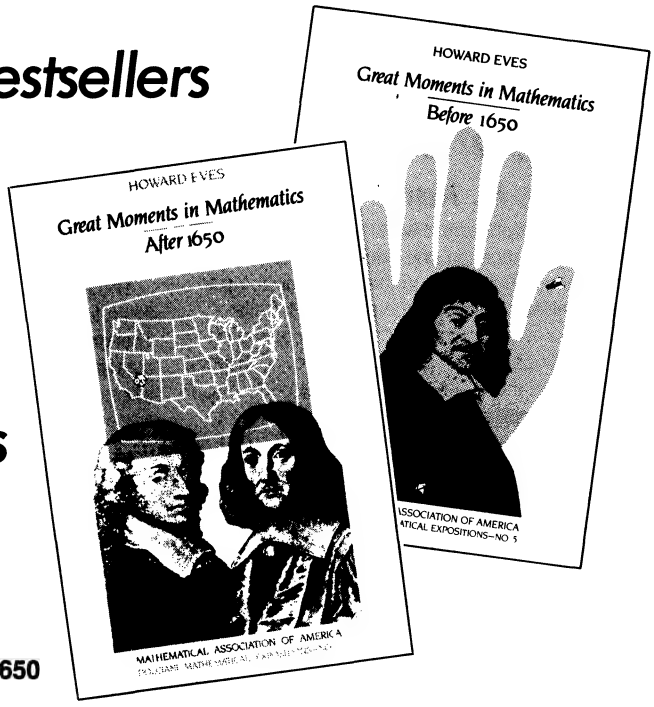
Princeton University Press
41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 258-4900
ORDERS: 800-PRS-ISBN (777-4726)

When was the last time a computer program helped you think about mathematics?

Call or write for our latest software catalog.
For PC or Macintosh.

Lascaux Graphics 3220 Steuben Ave. Bronx, NY 10467 (212) 654-7429

The MAA Bestsellers on the History of Mathematics



GREAT MOMENTS IN MATHEMATICS BEFORE 1650

Howard Eves

This outstanding volume (along with its companion Volume 7) has become an MAA bestseller. It contains 23 fascinating mathematical episodes from antiquity to about 1650.

270 pp., Paper, 1982, ISBN 0-88385-310-8

List: $17.00 MAA Member: $13.00

Catalog Number DEP-05

GREAT MOMENTS IN MATHEMATICS AFTER 1650

Howard Eves

This is a companion volume to Volume 5.

270 pp., Paper, 1982, ISBN 0-88385-311-6

Paper List: $17.00 MAA Member: $13.00
Catalog Number DEP-07 (Paper)

Hardbound List : $27.00
MAA Member: $20.00
Catalog Number DOL-07 (Hardbound)

Both books available in paperbound editions as a package: List: $31.50 MAA Member: $23.00

Catalog Number DEP-057

Howard Eves made a valuable contribution to the Dolciani Mathematical Exposition series . . . The twenty lectures included are a delight to read. They place each "great moment" in its historical context and lay special emphasis on human aspects of each achievement. No algebra or geometry teacher should be without this book.

Tom Walsh in *The Mathematics Teacher*

He has the knack of bringing ancient mathematics alive, at the same time relating it to modern ideas and developments. His work is readable and informative and indeed bears reading a second time. His exposition is clear and he is not content with generalities, but gives actual examples of early methods.

E.H. Lockwood in
The Mathematical Gazette

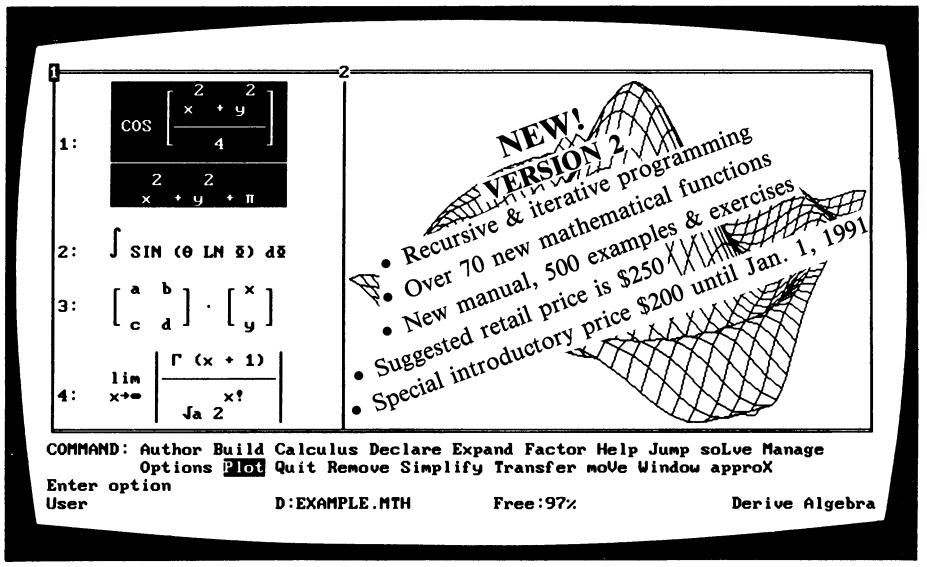
ORDER FROM

The Mathematical Association of America

1529 Eighteenth Street, N.W.
Washington, D.C. 20036

DERIVE

A Mathematical Assistant



2000 years of mathematical knowledge on a disk

DERIVE, **the successor to muMATH, is a powerful computer algebra system for your PC compatible computer that provides the following capabilities:**

- **Exact and approximate arithmetic to thousands of digits**
- **Equations, complex numbers, trigonometry, calculus, vectors, and matrices**
- **2D and 3D function plotting with zooming capability**
- **MDA, CGA, EGA, VGA, and Hercules graphics and text support**
- **Attractive 2D mathematical display of formulas**
- **Easy to use menu-driven interface with on-line help**
- **Ideal for engineers, scientists, students and teachers**
- **$200 plus shipping: Call or write for information.**

System requirements: IBM PC or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (360K) or a 3½ inch (760K) diskette drive. **Or** NEC PC-9801 or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (640K) diskette drive.

DERIVE and muMATH are trademarks of Soft Warehouse, Inc. Hercules is a trademark of Hercules Computer Technology, Inc. IBM is a registered trademark of International Business Machines Corp. MS-DOS is a registered trademark of Microsoft Corp. NEC is a registered trademark of Nippon Electric Company.

Soft Warehouse INC
3615 Harding Avenue, Suite 505 • Honolulu, HI 96816
(808) 734-5801 after noon PST
Handcrafted software for the mind.

©1988 Soft Warehouse, Inc.

Winner of the First MAA Book Prize . . .



Carus Mathematical Monograph, #19

Field Theory and its Classical Problems,

by Charles R. Hadlock

323 + xvi pp. Hardbound

List: $28.00 MAA Member: $22.00

Field Theory and its Classical Problems is one of those rare monographs that will capture and hold a mathematically prepared reader (one reasonably familiar with basic calculus and linear algebra) until the last page. Mathematicians will be intrigued by the development which places Galois Theory in an historical perspective and allows it to unfold from the emergence of the classical construction problems to the discovery of their solutions. The book presents some of history's greatest and most elegant mathematics.

Read what some reviewers say about this monograph:

"Hadlock has produced a pedagogic masterpiece . . . His idea of beginning with the three Greek problems and letting them fire the reader's spark plugs is brilliant . . . (his) ability to inject vitality and enthusiasm into mathematical text is remarkable. (George Piranian, University of Michigan)

"In the preface, the author claims to have written the book for himself . . . Fortunately for us, he chose to share his work with the mathematical community. I suggested the book as collateral reading in a one-semester course in Galois Theory and the students found it very readable and most helpful in establishing a motivation and direction . . . This is a delightful book for both student and teacher." (John D. Leadley, ***Mathematical Reviews)***

Table of Contents

Chapter 1—The Three Greek Problems
Constructible Lengths; Doubling the Cube; Trisecting the Angle; Squaring the Circle; Polynomials and Their Roots; Symmetric Functions; the Transcendence of π.

Chapter 2—Field Extensions
Arithmetic of Polynomials; Simple, Multiple, and Finite Extensions; Geometric Constructions Revisited; Roots of Complex Numbers; Constructibility of Regular Polygons I; Congruences; Constructibility of Regular Polygons II.

Chapter 3—Solution by Radicals
Statement of the Problem; Automorphisms and Groups; The Group of an Extension; Two Fundamental Theorems, Galois' Theorem; Abel's Theorem; Some Solvable Equations.

Chapter 4—Polynomials With Symmetric Groups
Background Information; Hilbert's Irreducibility Theorem; Existence of Polynomials over Q with Group S_n.



Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

HBJ Brings you Range, Depth, and Innovation in 1991

INTRODUCTORY ALGEBRA
and
INTERMEDIATE ALGEBRA
RONALD HATTON and GENE R. SELLERS
both of Sacramento City College

Now in Hardcover!
ELEMENTARY ALGEBRA
Second Edition
and
INTERMEDIATE ALGEBRA
Second Edition
YOSHIKO YAMATO and
MARY JANE CORDON
both of Pasadena City College

FINITE MATHEMATICS
WILLIAM OWEN and FRED CUTLIP
both of Central Washington University

ALGEBRA FOR COLLEGE STUDENTS
Third Edition
BERNARD KOLMAN, *Drexel University*
and ARNOLD SHAPIRO

ELEMENTARY LINEAR ALGEBRA WITH APPLICATIONS
Second Edition
RICHARD O. HILL, JR.,
Michigan State University

CALCULATOR ENHANCEMENT FOR A COURSE IN LINEAR ALGEBRA
DON LA TORRE, *Clemson University*

CALCULUS
One and Several Variables
ROBERT ELLIS and DENNY GULICK
both of University of Maryland
Just $29.95 (Net)

SINGLE-VARIABLE CALCULUS WITH DISCRETE MATHEMATICS
JOHN A. FEROE and
CHARLES STEINHORN
both of Vassar College

CALCULATOR ENHANCEMENT FOR A COURSE IN SINGLE VARIABLE CALCULUS
JOHN KENNELLY and
JAMES H. NICHOLSON
both of Clemson University

CALCULATOR ENHANCEMENT FOR A COURSE IN MULTIVARIABLE CALCULUS
JAMES A. RENEKE, *Clemson University*

CALCULATOR ENHANCEMENT FOR A COURSE IN DIFFERENTIAL EQUATIONS
THOMAS G. PROCTOR,
Clemson University

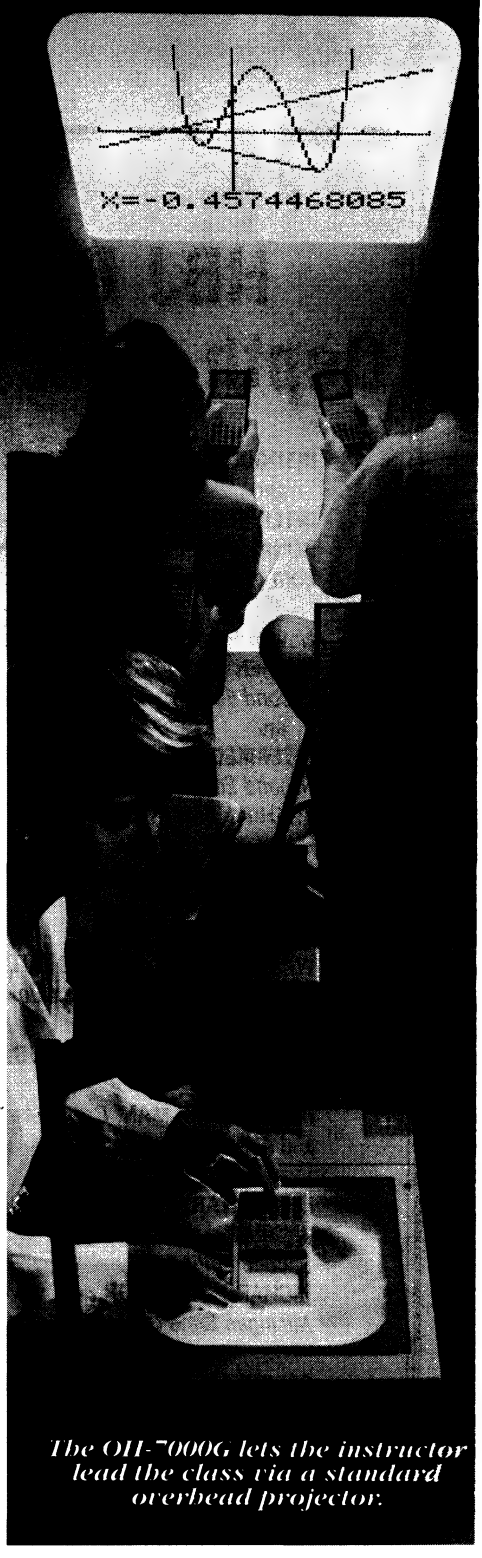
CALCULATOR ENHANCEMENT FOR A COURSE IN INTRODUCTORY STATISTICS
IRIS FETTA, *Clemson University*

ABSTRACT ALGEBRA
DENNIS KLETZING, *Stetson University*



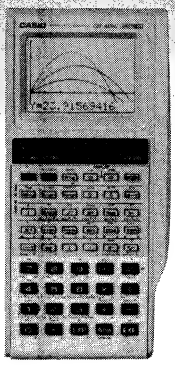
Harcourt Brace Jovanovich, Inc.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648
(708) 647-8822

The Power to Project Your Ideas



The OH-7000G lets the instructor lead the class via a standard overhead projector.

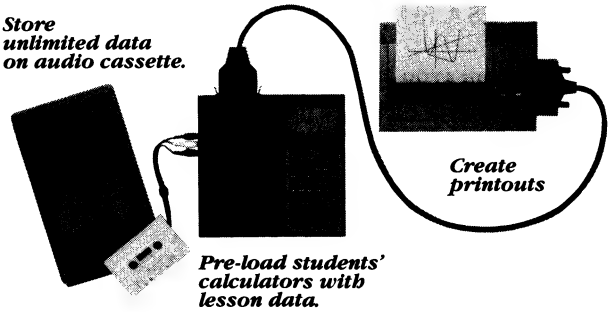
Casio has taken one of its most powerful graphic calculators and turned it into the most powerful **teaching tool** of its kind. The OH-7000G features a transparent, projectable screen so that teachers can lead their class just by using a standard overhead projector.



OH-7000G

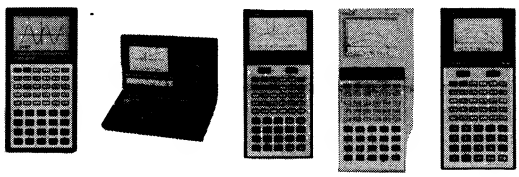
Since inventing the graphic calculator in 1985, Casio has created an entire family of exciting classroom products. Five different graphic calculator models are available, each offering **hundreds of powerful, built-in functions**.

With our **printers**, you can create "hard copy" of graphs and calculations for photocopying onto test papers or transparencies. The **FA-80 Data Interface** lets you store unlimited data on audio cassettes.



The possibilities are endless. That's why the **NCTM curriculum** calls for the use of graphic calculators. Many **new textbooks** are being published especially for use with this technology.

So, take a bold, new step and enter Casio's "Electronic Classroom" – a place where the solution is always clear.



CASIO

Creators of th Graphic Calculato

570 Mt. Pleasant Avenue, Dover, New Jersey 07

For more information about Casio's family of graphic calculators, and to place an order, call one of the educational distributors listed below.

ABM
800-234-3396
(In TX 214-869-0721)
(In MO 816-861-1113)
(In CA 213-946-6680)

Allied Communications
800-843-3660
(In PA 800-843-3660)

Allied National
800-999-8099
(In MI 313-543-1232)

Calculators Inc.
800-533-9921
(In MN 800-533-9921)

Carolina Wholesale
800-521-4600
(In NC 704-598-8101)

Colborn School Supply
800-548-7031
(In CO 303-778-1220)
(In MT 406-245-3158)

Copco
800-446-7021
(In OH 800-589-3006)

Dale Seymour Publications
800-872-1100
(In CA 800-222-0766)

Douglas Stewart
800-356-8370
(In WI 608-221-1155)

E.A.I.
800-272-0272
(In NJ 201-891-9466)

El Dorado Trading
800-227-8292
(In CA 415-494-1995)

Entertainment Marketing
800-347-2001
(In TX 713-995-4433)

Global Products
800-633-0633
(In IL 708-397-4944)

Longino Distributors
800-633-6224
(In NC 704-873-3282)

NASCO
800-558-9595
(In WI 414-563-2446)

Nemesis Distributing
800-940-7407
(In FL 305-477-8822)

Servco Pacific
(In HI 808-841-7566)

Tam's
800-421-5188
(In CA 800-244-5624)

Taylor Electric
800-558-6970
(In WI 800-242-8940)

Techline
800-777-3635
(In VA 703-389-0857)

Valley Business Machines
(In UT 801-969-6303)

Vistatech
800-847-9851
(In NY 212-254-9851)
(In CA 213-602-0277)

CASIO®

Mathematics Graduate Studies at The Ohio State University

The Department of Mathematics at Ohio State offers several options for graduate study: a broad Ph.D. degree program with specialization in nearly all branches of contemporary mathematics, pure and applied; an M.S. degree program with options in pure or applied mathematics, an elective focus in Actuarial Science, and a dual M.S. degree program with Computer Science; a content oriented M.A. degree program designed for secondary level teachers of mathematics. MS/PhD applicants must present scores from the GRE Advanced Subject test in Mathematics and should have at least a one year background in advanced calculus and linear/abstract algebra.

Students are encouraged to begin their studies at Ohio State in June, 1991. A number of Summer Fellowships providing stipends ranging from $1400 – $2000 in addition to a waiver of all tuition and fees are anticipated. Academic year financial support is available in the form of Teaching Associateships and Fellowships with stipends ranging from $900 to $1000 per month in addition to a waiver of all tuition and fees. Academic year fellowship recipients may request halftime teaching positions offering additional monthly stipends of $460. M.A. applicants may request double teaching loads with combined monthly stipends of $1840. Summer support in the form of teaching and research associateships are also available for many continuing students.

For 1990-91, the Department anticipates being able to offer significant supplemental funding to highly qualified individuals from groups (Blacks, Hispanics, and Women) which are currently underrepresented in the mathematics community.

For further information write to:
Professor Harry Allen Ohio State University 231 West 18th Avenue Columbus Ohio 43210

The Ohio State University
invites applications for

TRANSIT

An NSF project to train **School/University Teams** that will establish regional technology centers. Regional team training will be provided through summer in-service sessions and academic year follow-up conferences at Ohio State University. Local living expenses with some stipend and travel support is available. The regional centers, in turn, will train teachers as school technology specialists at their regional site. Partial first year support from Ohio State University will be available for a few regional sites.

Deadline for completed applications is February 15, 1991.

Write TRANSIT, c/o Frank Demana & Bert Waits, The Ohio State University, Mathematics Department, 231 West 18th Avenue, Columbus, OH 43210.

STUDIES IN THE HISTORY OF MATHEMATICS

STUDIES IN THE HISTORY OF MATHEMATICS

Esther R. Phillips, Editor

Esther Phillips has brought together a collection of articles showing the sweep of recent scholarship in the history of mathematics. The material covers a wide range of current research topics: algebraic number theory, geometry, topology, logic, the relationship between mathematics and computing, partial differential equations, and algebraic geometry.

320 pp., 1987, ISBN 0-88385-128-8

List: 36.50 MAA Member: $28.00

Catalog Number MAS-26

This is an excellent book! It is a very interesting and exciting book to read. The author does an extremely nice job of bringing together most, if not all, the mathematicians that were involved in a particular area of mathematics. The sources listed at the end of each section give the reader an opportunity to look up other resources pertaining to the particular subjects, a feature that is definitely lacking in many history books. The content of the book is choice. The professional mathematician would definitely want to have a copy of this book.

Barney Erikson in *The Mathematics Teacher*

Order from: The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200



Differential Equations: A Modeling Approach
Frank R. Giordano and Maurice D. Weir, *U.S. Military Academy and U.S. Naval Postgraduate School, respectively*

Now your students can clearly see how differential equations apply to realistic problems. This new text by Giordano and Weir focuses on mathematical modeling, showing students a natural progression of related ideas, and offering motivation to take the next step along the way. Unlike other texts that use a "solutions technique" approach, **Differential Equations: A Modeling Approach** puts into perspective how the various solutions techniques students learn actually fit together — the "why" as well as the "how-to."

Discrete Algorithmic Mathematics
Stephen B. Maurer and Anthony Ralston, *Swarthmore College and SUNY-Buffalo, respectively*

Turn to Maurer and Ralston for discrete mathematics with a basic difference: a unique algorithmic approach. The authors present discrete mathematics as a unified topic, by viewing it through an algorithmic "prism." A wide range of applications, flexible coverage of core discrete math topics, and extensive coverage of linear algebra also add to this text's substantial value.

Addison-Wesley
1 Jacob Way • Reading, MA • 01867
(617) 944-3700

From the publisher that's serious about math

▼
ALSO NEW FOR '91!

Developmental Mathematics

Basic Mathematics, Sixth Edition
Introductory Algebra, Sixth Edition
Intermediate Algebra, Sixth Edition
Mervin L. Keedy and Marvin L. Bittinger, *Purdue University and Indiana University-Purdue University at Indianapolis, respectively*

Concepts and Applications of Intermediate Algebra, Second Edition
Marvin L. Bittinger, Mervin L. Keedy, and David Ellenbogen, *Indiana University-Purdue University at Indianapolis, Purdue University, and St. Michael's College, respectively*

Elementary Algebra
Intermediate Algebra
Mark Dugopolski, *Southeastern Louisiana University*

Statistics

Introductory Statistics, Third Edition
Neil A. Weiss and Matthew Hassett, *both of Arizona State University*

Statistics Methods for Business and Economics, Fourth Edition
Donald L. Harnett and Ashok Soni, *both of Indiana University*

▼
OTHER TITLES OF INTEREST

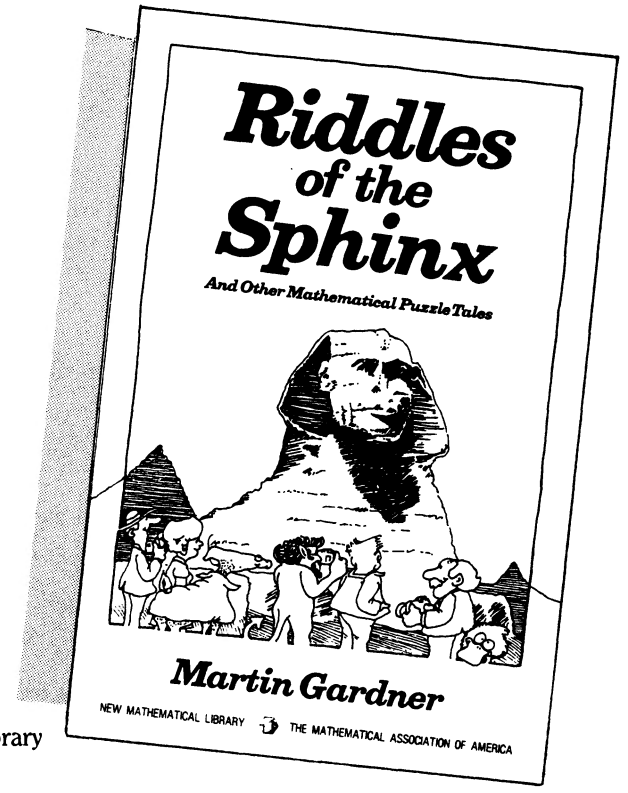
Precalculus: Functions and Graphs
A. Robert Marshall, *Cuesta College*

Mathematics for Elementary School Teachers: A Problem Solving Approach
Joseph Newmark, *CUNY College of Staten Island*

Calculus: A Complete Course
Robert A. Adams, *University of British Columbia*

Riddles of the Sphinx

and other mathematical puzzle tales.



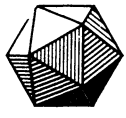
by Martin Gardner
Volume 32 in the New Mathematical Library
184 pp., Paper. ISBN-0-88285-632-8
List: $14.50 MAA Member: $12.50

Martin Gardner has charmed readers for over fifty years with his delightful books and articles on science. He is best known for the popular column, *"Mathematical Games,"* which appeared in *Scientific American* for twenty-five years. Generations of scientists and mathematicians have been inspired by his writing and the MAA is proud to include his name in its list of authors.

This book was drawn from Gardner's column in *Isaac Asimov's Science Fiction Magazine.* The riddles presented here incorporate the responses of his initial readers, along with additions suggested by the editors of the New Mathematical Library. Each chapter (riddle) poses a problem answered in the First Answers section. The solution in turn raises another problem that is solved in the Second Answers section. This may suggest a third question and in several instances there is a fourth. Gardner draws us from questions to answers always presenting us with new riddles—some as yet unanswered. There are 125 different pieces altogether.

Solving these riddles is not simply a matter of logic and calculation, although these play a role. Luck and inspiration are factors as well, so beginners and experts alike may profitably exercise their wits on Gardner's problems, whose subjects range from geometry to word play to questions relating to physics and geology.

We guarantee that you will solve some of the riddles, be stumped by others, and be amused by almost all of the stories and settings that Gardner has devised to raise these questions.



Order from:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA